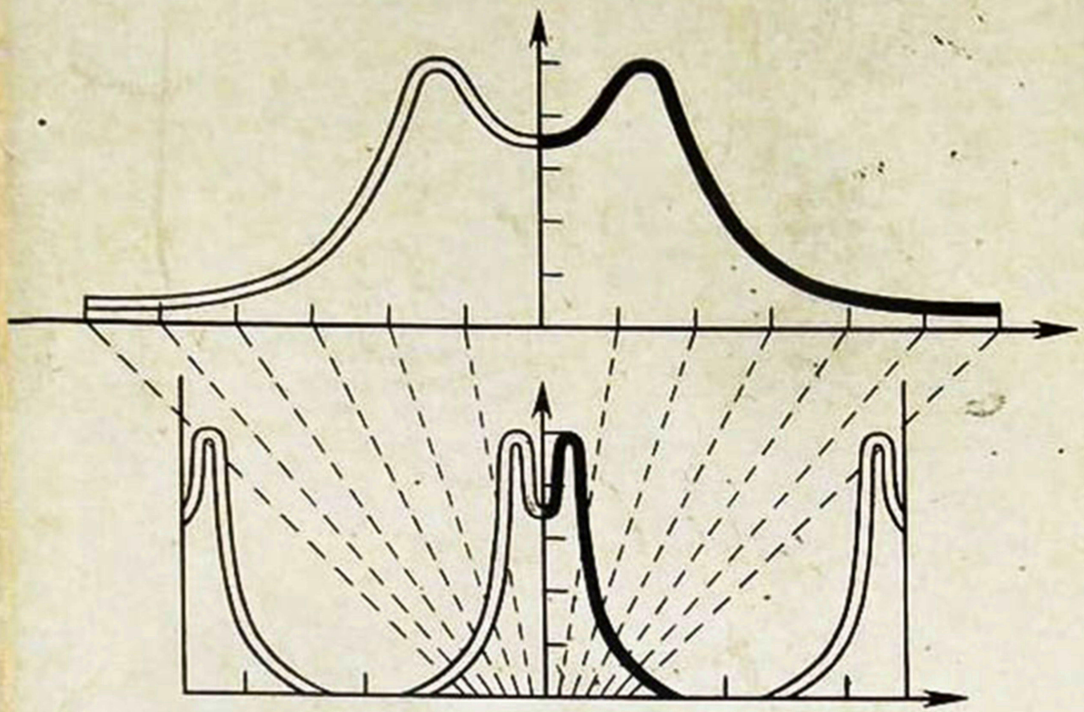
RADIO CIRCUITS AND SIGNALS

I. S. GONOROVSKY



MIR
PUBLISHERS
MOSCOW

RADIO CIRCUITS AND SIGNALS

I. S. GONOROVSKY

I.S. GONOROVSKY

RADIO CIRCUITS AND SIGNALS

И. С. ГОНОРОВСКИЙ

РАДИОТЕХНИЧЕСКИЕ ЦЕПИ И СИГНАЛЫ

«Советское Радио»
Москва

RADIO CIRCUITS AND SIGNALS

I.S. GONOROVSKY

Translated from the Russian
by
N. UTKIN

MIR PUBLISHERS
MOSCOW

First published 1981
Revised from the 3rd 1977 Russian edition

The Greek Alphabet

A α	Alpha	I ι	Iota	P ρ	Rho
B β	Beta	K κ	Kappa	Σ σ	Sigma
Γ γ	Gamma	Λ λ	Lambda	T τ	Tau
Δ δ	Delta	M μ	Mu	Υ υ	Upsilon
E ε	Epsilon	ν	Nu	Φ φ	Phi
Z ζ	Zeta	Ξ ξ	Xi	X χ	Chi
H η	Eta	O o	Omicron	Ψ ψ	Psi
Θ θ	Theta	Π π	Pi	Ω ω	Omega

На английском языке

© Советское Радио, 1977
© English translation, Mir Publishers, 1981

CONTENTS

Preface 11

Chapter 1. Introduction 13

1.1. Basic Applications of Radio Engineering 13
1.2. Transmission of Signals. Radio Frequency Bands and Propagation of Radio Waves 14
1.3. Fundamental Radio Processes 18
1.4. Analog, Discrete and Digital Signals and Circuits 21
1.5. Methods of Analysis of Radio Circuits 23
1.6. Noise Immunity of Communication Channels 29
1.7. Objectives of the Present Course 31

Chapter 2. Signals 32

2.1. General 32
2.2. Expansion of an Arbitrary Signal in a Given Set of Functions 34
2.3. Harmonic Analysis of Periodic Oscillations 38
2.4. Spectra of Simple Periodic Oscillations 42
2.5. Power Distribution in the Spectrum of a Periodic Oscillation 47
2.6. Harmonic Analysis of Nonperiodic Oscillations 48
2.7. Some Properties of Fourier Transforms 52
2.8. Energy Distribution in the Spectrum of a Nonperiodic Oscillation 58
2.9. Examples of Determining Spectra of Nonperiodic Oscillations 58
2.10. Relation Between the Duration of a Signal and the Width of Its Spectrum 66
2.11. Infinitely Short Pulse of Unit Area (Delta Function) 68
2.12. Spectra of Some Nonintegrable Functions 71
2.13. Representation of Signals on the Plane of a Complex Variable 76
2.14. Expansion of Band-Limited Signals in Functions of the Form sinc (x). 84
2.15. Sampling Theorem in the Frequency Domain 88
2.16. Correlation Analysis of Deterministic Signals 90
2.17. Relationship Between the Autocorrelation Function and the Spectral Characteristic of a Signal 95
2.18. Coherence. 95

Chapter 3. Radio Signals 98

3.1. General 98
3.2. Amplitude-Modulated Radio Signals 100
3.3. Frequency Spectrum of an Amplitude-Modulated Signal 102
3.4. Angle Modulation. Phase and Instantaneous Frequency of a Wave 109
3.5. Frequency Spectrum of an Angle-Modulated Wave. General Relations 114
3.6. Wave Spectrum in Harmonic Angle Modulation 115
3.7. Spectrum of a Radio Pulse with a Frequency-Modulated Carrier 120
3.8. Spectrum of an Amplitude-Frequency Modulated Wave . . 123
3.9. Envelope, Phase and Frequency of a Narrow-Band Signal 127
3.10. Analytic Signal 133
3.11. Autocorrelation Function of a Modulated Wave 138
3.12. Sampling of a Narrow-Band Signal 141

Chapter 4. Main Characteristics of Random Signals 146

4.1. Random Processes. General 146
4.2. Types of Random Processes. Examples 150
4.3. Power Spectral Density of a Random Process 158
4.4. Relationship Between the Energy Spectrum and the Autocorrelation Function of a Random Process 160
4.5. Cross-Correlation Function and Cross-Spectral Density of Two Random Processes 165
4.6. Narrow-Band Random Process 168
4.7. Oscillation Modulated in Amplitude by a Random Process 177
4.8. Oscillation Modulated in Phase by a Random Process. Probability Density 181

Chapter 5. Linear Radio Circuits with Constant Parameters 185

5.1. General 185
5.2. Fundamental Properties of Active Networks. Definitions 185
5.3. Active Two-Port Network as a Linear Amplifier 192
5.4. Transistor Amplifier 195
5.5. Electron-Tube Amplifier 202
5.6. Aperiodic Amplifier 204
5.7. Resonance Amplifier 207
5.8. Feedback in an Active Two-Port Network 210
5.9. Use of Negative Feedback for Improving Amplifier Characteristics 215
5.10. Stability of Linear Active Circuits with Feedback. Algebraic Stability Criterion 222
5.11. Frequency Stability Criteria 227

Chapter 6. Transmission of Deterministic Oscillations Through Linear Circuits with Constant Parameters 232

6.1. General 232
6.2. Spectral Method 233
6.3. Superposition Integral Method 234

6.4. Transmission of Discrete Signals Through an Aperiodic Amplifier 236
6.5. Differentiation and Integration of Signals 239
6.6. Analysis of Radio Signals in Selective Circuits. Approximate Spectral Method 246
6.7. Simplification of the Superposition Integral Method (Envelope Method) 249
6.8. Transmission of a Radio Pulse Through a Resonance Amplifier 250
6.9. Linear Distortion of Oscillations with Continuous Amplitude Modulation 256
6.10. Transmission of a Phase-Shift Keyed Oscillation Through a Resonance Circuit 261
6.11. Transmission of a Frequency-Shift Keyed Oscillation Through a Selective Circuit 263
6.12. Transmission of a Frequency-Modulated Oscillation Through a Selective Circuit 269

Chapter 7. Transmission of Random Oscillations Through Linear Circuits with Constant Parameters 275

7.1. Conversion of Characteristics of a Random Process 275
7.2. Characteristics of Intrinsic Noise in Radio Circuits 277
7.3. Differentiation of a Random Function 284
7.4. Integration of a Random Function 287
7.5. Normalization of Random Processes in Narrow-Band Linear Circuits 289
7.6. Distribution of a Sum of Harmonic Oscillations with Random Phases 291

Chapter 8. Nonlinear Circuits and Methods of Their Analysis 296

8.1. Nonlinear Elements 296
8.2. Approximation of Nonlinear Characteristics 299
8.3. Action of Harmonic Oscillations on Circuits with Lag-Free Nonlinear Elements 302
8.4. Nonlinear Resonance Amplification 309
8.5. Frequency Multiplication 313
8.6. Amplitude Limiting 315
8.7. Nonlinear Rectifying Circuit 320
8.8. Amplitude Detection 324
8.9. Frequency and Phase Detection 331
8.10. Conversion of the Frequency of a Signal 337
8.11. Synchronous Detection 341
8.12. Generation of Amplitude-Modulated Oscillations 341

Chapter 9. Harmonic Self-Oscillators 345

9.1. Self-Oscillating System 345
9.2. Origination of Oscillations in a Self-Oscillator 348
9.3. Steady-State Operation of a Self-Oscillator. Phase Balance 352
9.4. Soft and Hard Self-Excitation Conditions 356
9.5. Examples of Self-Oscillator Circuits 358
9.6. Nonlinear Equation of the Self-Oscillator 361

9.7. Approximate Solution of the Nonlinear Equation of the Self-Oscillator 364
9.8. Self-Oscillators with Internal Feedback 367
9.9. Self-Oscillator with a Delay Line in the Feedback Circuit 370
9.10. Action of a Harmonic E.M.F. on Circuits with Positive Feedback. Regeneration 373
9.11. Action of a Harmonic E.M.F. on a Self-Oscillator. Frequency Pulling 376
9.12. Angle Modulation in the Self-Oscillator 380
9.13. *R-C* Oscillators 382

Chapter 10. Circuits with Variable Parameters 388

10.1. General 388
10.2. Transmission of Oscillations Through Linear Circuits with Variable Parameters. Transfer Function 391
10.3. Modulation as a Parametric Process 394
10.4. Impulse Response of a Parametric Circuit 396
10.5. Energy Relations in a Circuit with a Nonlinear Reactive Element in the Case of Biharmonic Excitation 401
10.6. Principle of Parametric Amplification of Oscillations . . . 408
10.7. Equivalent Circuit of a Harmonically Varying Capacitance or Inductance 410
10.8. Single-Tuned Parametric Amplifier 414
10.9. Double-Frequency Parametric Amplifier 416
10.10. Frequency Conversion by Means of a Nonlinear Reactive Element 422
10.11. Free Oscillations in a Circuit with Periodically Varying Capacitance 423
10.12. Parametric Oscillators 431

Chapter 11. Effect of Random Oscillations on Nonlinear and Parametric Circuits 434

11.1. General 434
11.2. Conversion of a Random Process in Lag-Free Nonlinear Circuits 434
11.3. Conversion of Energy Spectrum in a Lag-Free Nonlinear Element 438
11.4. Effect of Narrow-Band Noise on an Amplitude Detector . . . 441
11.5. Joint Effect of a Harmonic Oscillation and Normal Noise on an Amplitude Detector 445
11.6. Joint Effect of a Harmonic Oscillation and Normal Noise on a Frequency Detector 449
11.7. Interaction of a Harmonic Oscillation and Normal Noise in an Amplitude Limiter with a Resonant Load 453
11.8. Autocorrelation Function and Energy Spectrum of a Random Process in a Parametric Circuit 457
11.9. Effect of Multiplicative Noise on the Distribution Law of a Signal 461

Chapter 12. Matched Filtering of Signals on a Noisy Background . . 466

12.1. General 466
12.2. Matched Filtering of a Given Signal 467
12.3. Impulse Response of the Matched Filter. Physical Realizability 472
12.4. Signal and Noise at the Output of the Matched Filter . . . 475
12.5. Examples of Constructing Matched Filters 477
12.6. Shaping of the Signal which is the Complex Conjugate of a Given Signal 489
12.7. Matched Filtering of a Specified Signal in the case of Non-white Noise 492
12.8. Filtering of a Signal of Unknown Epoch Angle 493
12.9. Matched Filtering of a Complex Signal 495

Chapter 13. Discrete Processing of Signals. Digital Filters 500

13.1. General 500
13.2. Algorithm of Discrete Convolution (in the Time Domain) 502
13.3. Discrete Fourier Transforms 505
13.4. Error of Sampling Signals of Finite Length 510
13.5. Discrete Laplace Transforms 513
13.6. Transfer Function of a Discrete Filter 515
13.7. Transfer Function of a Recoursive Filter 519
13.8. Use of the z-Transform Method for the Analysis of Discrete Signals and Circuits 523
13.9. z-Transform of Time Functions 524
13.10. z-Transforms of Transfer Functions of Discrete Circuits 527
13.11. Examples of the Analysis of Discrete Filters, Based on the z-Transform Method 530
13.12. Analog-to-Digital Conversion. Quantization Noise 536
13.13. Digital-to-Analog Conversion and Restoration of a Continuous Signal 541
13.14. Speed of Response of the Arithmetic Unit of a Digital Filter. Rounding-Off Noise 545

Chapter 14. Representation of Oscillations by Some Special Functions 548

14.1. General 548
14.2. Orthogonal Polynomials and Continuous Functions 549
14.3. Examples of Application of Continuous Functions 555
14.4. Definition of the Walsh Functions 559
14.5. Examples of Application of the Walsh Functions 569
14.6. Mutual Spectrum of Two Different Complete Orthonormal Sets of Functions 575
14.7. Discrete Walsh Functions 577

Chapter 15. Elements of Synthesis of Linear Radio Circuits 582

15.1. General 582
15.2. Some Properties of the Transfer Function of a Two-Port Network 583
15.3. Relation Between the Amplitude-Frequency and Phase-Frequency Characteristics of a Two-Port Network 584

15.4. Representation of a General Two-Port Network by Cascade Connection of Elementary Two-Port Networks 589
15.5. Realization of a Typical Second-Order Link 591
15.6. Realization of a Phase-Correcting Circuit 594
15.7. Synthesis of a Two-Port Network by the Given Amplitude-Frequency Characteristic 598
15.8. Synthesis of a Low-Pass Filter. Butterworth Filter 600
15.9. Chebyshev Low-Pass Filter 605
15.10. Synthesis of Various Filters on the Basis of a Low-Pass Filter 608
15.11. Sensitivity of Circuit Characteristics to Changes in the Parameters of the Circuit Elements 609
15.12. Simulation of Inductance by Means of an Active *R-C* Circuit. Gyrator 612
15.13. Some Specific Features of the Synthesis of Digital Filters 616

Appendix I. Signal with Minimum Product of Duration and Frequency Band 622

Appendix II. Autocorrelation Function of a Signal on the Time-Frequency Plane 625

Literature 629

Symbols 632

Index 635

PREFACE

Since 1971, when this textbook was last published, radioelectronics has made great progress, necessitating our extensive revision and supplementing of the material, although the same guiding principles have been followed as in the first two editions.

The wide use of discrete and digital radioelectronic circuits has made it impossible to restrict the course on radio circuits and signals to *analog* circuits and signals alone, while the development of integrated microcircuits, based on an extensive application of circuit *synthesis* methods, has prompted us to expand it beyond the study of the techniques of circuit *analysis*, and, finally, the ever increasing introduction of statistical methods in all branches of radio engineering and electronics has required a more comprehensive study of random signals and their transformation in radio circuits.

This edition, therefore, contains five entirely new chapters — Main Characteristics of Random Signals (Ch. 4), Transmission of Random Oscillations Through Linear Circuits with Constant Parameters (Ch. 7), Discrete Processing of Signals and Digital Filters (Ch. 13), Representation of Oscillations by Some Special Functions (Ch. 14), and Elements of Synthesis of Linear Radio Circuits (Ch. 15).

Chapter 5 of the previous edition, dealing with the theory of active linear circuits with feedback loops, has been totally rewritten, and methodological changes have been made in the rest of the material on the basis of criticisms and numerous suggestions made by college teachers and radio specialists.

Seeking to inspire students to do independent research work, we have expanded the scope of the book to include, along with the subject matter for compulsory study, some additional, more complex material for advanced students. This material is printed in small type.

The book is so arranged that minor abridgements, which may be necessitated by the inadequate theoretical knowledge of the students, will not impair integrity of the presentation.

In conclusion we should like to thank Professor N. N. Fedorov, Senior Readers S. I. Baskakov, and I. V. Belousova, Assistant Lecturer V. I. Bogatkin, Senior Reader V. P. Zhukov, Senior Teacher N. N. Ivanova, Senior Readers V. G. Kartashev, A. M. Nikolaev, and B. P. Pollak, and Senior Teacher V. V. Shtykov of the Radio Engineering Department of the Moscow Power Institute for their careful reading and checking of the original manuscript. We are grateful to Senior Reader L. A. Chinenkov of the Communications Equipment Sub-Faculty of the Novosibirsk Institute of Communications for his numerous valuable comments on the previous edition of this book. Deep appreciation goes to the teaching and research staff of the Radio Engineering Sub-faculty of the Moscow Aviation Institute for their invaluable help in the preparation of the manuscript.

I. S. Gonorovsky

Chapter 1

INTRODUCTION

1.1. BASIC APPLICATIONS OF RADIO ENGINEERING

Modern radio engineering is a powerful tool in speeding up the scientific and technical progress. Radio engineering has penetrated all branches of the national economy, science, industry, our culture and everyday life.

One of the most important applications of radio engineering involves long-distance communication by means of electromagnetic waves. The development of various specialized branches of radio engineering is closely allied to the general use of radio for broadcasting and communication, while television covers steadily expanding regions in many parts of the globe. Radio equipment provides for reliable round-the-clock communication with marine vessels, aircraft, and spaceships. Radio engineering systems enable us to effect interplanetary communications and to provide remote control of apparatus used for exploration of other planets. Such branches of radio engineering as radio detection and ranging (or radar), radio navigation, radio telemetry, and radio control, which just a few years ago were regarded as new techniques, are now in general use.

However, the above applications by no way exhaust all the possibilities of modern radio engineering. Radio methods have penetrated into well-known sciences and led to their qualitative change and further development. New sciences have been born, such as radio physics and radio astronomy.

Radio techniques and methods are widely used in experimental physics, including nuclear physics, in instruments and apparatus measuring transient processes and various non-electrical quantities (pressure, mechanical vibration, small displacements and so on), in studying physical phenomena occurring in the ionosphere, and in time service.

The extensive use of radio methods for solving various problems not associated with radiation of electromagnetic waves has given rise to a novel science that embraces both radio engineering and electronics. This branch of science is commonly referred to as *radioelectronics.*

The ever wider use of high-speed electronic computers for calculations, control and data acquisition is a great achievement of radioelectronics. Cybernetic systems playing a decisive role in

process control and automation is one of the major areas of development of radioelectronics. Radioelectronic equipment is widely used in medical studies (for diagnostics) and in the manufacture of artificial organs or devices that are employed to compensate for partially or completely lost functions of a human organism.

From what it is said above we may draw a conclusion that applications of radioelectronics are many-sided and that its role in the future progress of mankind will incessantly gain in importance.

Since the very date of invention of the first radio receiver by A. S. Popov in 1895 and up to the present day the main application of radio engineering has been *transmission of information* by means of electromagnetic waves. This is a principal difference between radio engineering and electrical engineering, since the latter deals with transmission of energy (e.g., long-distance transmission of electric power over a high-voltage line) instead of information.

1.2. TRANSMISSION OF SIGNALS. RADIO FREQUENCY BANDS AND PROPAGATION OF RADIO WAVES

Thus, the transmission of a message over a distance is the basic task of radio engineering. Distance separates the transmitter and the receiver, the sender of control signals and the actuating mechanism, the process being studied and the measuring system, the source of cosmic radiation and the recorder of a radio telescope, and individual units of a computer. In other words, distance separates the source of information from its recipient.

The distances over which signals are transmitted may be either insignificant (transmission of commands from one unit of an electronic computer to another) or extremely large (intercontinental or cosmic communications). Messages can be transmitted over wire (cable) or waveguide lines or through free space. It is obvious that to transmit a signal, such physical processes should be used as possess the ability to propagate through space. These include electromagnetic oscillations (radio waves) that radio engineering makes use of. Any physical process employed as a carrier of information must be capable of taking the whole set of states that could be used to establish unambiguously the corresponding state of the object or process that provides the information to be transmitted, i.e., the source of information.

To this end, radio waves are subjected to *modulation*. The process of modulation consists in that a high-frequency oscillation capable of propagating over vast distances is given some properties characterizing the message to be transmitted. The oscillation is therefore used as a carrier of the message. For this purpose one or several parameters of the oscillation are varied according to the same law as governs the changes of the transmitted message. Depending

on the oscillation parameter varied (amplitude, frequency or phase), the following three fundamental types of modulation are distinguished: amplitude, frequency and phase modulation*.

The inverse transformation of electromagnetic waves into the initial signals at the receiving station is called *demodulation* or *detection* (amplitude, frequency or phase, respectively).

As a rule, modulation does not affect the ability of radio-frequency oscillations to propagate in free space; however, a proper choice of the wavelength (or the carrier frequency) is of vital importance for establishing reliable radio communication.

The choice of a frequency band for any particular radio communication system is governed by the following factors.

1. The character of propagation of electromagnetic waves in the given frequency band, depending on the season, time of day, atmospheric conditions, solar radiation, etc.

2. Technical capabilities: directional radiation, use of an appropriate antenna system, generation of high-power oscillations and their control (modulation), receiver design, etc.

3. Interference and noise characteristics in the given band.

4. The character of the message to be conveyed, that is the "spectrum width" of the modulating frequencies and the desired method of modulation (frequency, amplitude, etc.).

In practice those sub-bands are utilized that provide for the best conditions of propagation of radio waves and sufficiently satisfy all the other requirements stated above.

At present, radio engineers and scientists are investigating insufficiently known wavelengths with a view to mastering extremely low and high frequencies, including light waves. This is not so strange as it might appear at first sight, because radio and light waves are of the same nature: they are both electromagnetic waves.

The practical subdivision of electromagnetic waves into bands is given in Table 1.1.

Communication with the use of myriametric and kilometric waves, which was used in radiotelegraphy in the early days of radio engineering, suffers from the following serious disadvantages.

First, it is necessary to have a transmitter of a very high power output to compensate for the heavy absorption of the surface wave energy during its travel along the earth surface.

Second, these waves cannot be used for transmission of complex signals, because of the extremely high ratio of the signal spectrum width to the carrier frequency.

Hectometric waves are widely used in radio broadcasting. Radio waves longer than 1,000 m provide for stable reception, but they are difficult to utilize for long-distance communication, because

* There are also various types of pulse modulation, based on changing the parameters of a train of electric pulses.

Table 1.1

Waves	Wavelength range	Frequency range	Obsolete terms
Decamegametric	100,000 to 10,000 km	3 to 30 Hz	
Megametric	10,000 to 1,000 km	30 to 300 Hz	
Hectokilometric	1,000 to 100 km	300 to 3,000 Hz	
Myriametric	100 to 10 km	3 to 30 kHz	Super-long waves
Kilometric	10 to 1 km	30 to 300 kHz	Long waves
Hectometric	1,000 to 100 m	300 to 3,000 kHz	Medium waves
Decametric	100 to 10 m	3 to 30 MHz	Short waves
Metric	10 to 1 m	30 to 300 MHz	
Decimetric	100 to 10 cm	300 to 3,000 MHz	
Centimetric	10 to 1 cm	3 to 30 GHz	
Millimetric	10 to 1 mm	30 to 300 GHz	
Decimillimetric	1 to 0.1 mm	300 to 3,000 GHz	Submillimetric waves
Light	Less than 0.1 mm	Higher than 3 GHz	

Note. Wavelength λ is related to oscillation period T or frequency $f = 1/T$ by the relation $\lambda = cT = c/f$, where $c = 3 \times 10^8$ m/s is the velocity of propagation of electromagnetic waves in a vacuum.

of the considerable absorption of the surface wave energy. Therefore, these waves are mainly employed for local broadcasting within zones several hundred kilometres in radius. Only a few stations are known which have super-high power output and cover large areas. The USSR with its vast territory has the most powerful radio stations operating in this frequency range.

The decametric waves are advantageous in that long distances can be covered with a transmitter of comparatively low power output, and directional radiation is possible. The main disadvantage is the variation of the level of the received signal (fading), often accompanied by heavy distortion of complex signals consisting of a large number of components having different frequencies. Interference conditions, which depend on frequency, can be different for different components of the signal spectrum. This phenomenon, known as selective fading, results in a temporary dropout of individual components from the spectrum of the signal or, on the contrary, in the amplification of the amplitudes of these components. Thus, at the receiving end, the correct relation between individual spectrum components of the signal becomes disturbed and its tone and intelligibility distorted. The wider the signal spectrum, the stronger the effect of selective fading, therefore such complex signals as those used in television practically cannot be transmitted on decametric waves.

The large amount of experimental data obtained on the propagation of decametric waves has made it possible to determine the optimum wavelengths to suit the diurnal and seasonal variations in the propagation conditions, this having opened the way to the extremely wide application of short-wave broadcasting.

The mastering of metric (VHF) waves has resulted in the development of new branches of broadcasting — telecasting in particular. The metric band provides an advantageous combination of two factors: the use of a very high radiation frequency makes it possible to broaden the frequency band of the message being conveyed, because the conditions of transmission and amplification of signals in radio equipment are mainly defined by the relative width of the signal spectrum, while the propagation of metric waves within the direct visibility range almost completely eliminates the distortion of the signal due to the interference of waves spreading along different paths.

Of course, the fact that regular reception on metric waves is possible only within the direct visibility range is a substantial disadvantage. The communication range is usually increased by using very high antennas. The last decades have seen the development of what is known as radio relay systems consisting of a chain of metric-wave transceiver stations positioned along a communication line at a distance of several dozens of kilometres from one another. Such lines make it possible to effect multichannel communication and exchange of TV programs between fairly distant points.

Millimetric and shorter waves are now finding an ever wider practical application.

From the above review it is clear that the development of radio engineering is characterized by the continuous expansion of the frequency bands in use.

It is known from physics that effective radiation of electromagnetic energy can only be carried out if the geometrical dimensions of the antenna system are commensurable with the wavelength used. Therefore, very long waves are difficult to utilize. On the contrary, the use of light waves enables one to obtain small-size radiation devices with very sharp directivity and tremendous concentration of energy in the beam. For example, a light beam sent from the Earth forms on the Moon's surface a spot merely a few hundred metres in diameter. However, the transmission of information through the use of light waves is hindered by the difficulties associated with modulation, reception, weather effects, etc.

Lately, a number of promising trends have become apparent, pointing to the possibility that the drawbacks resulting from the electromagnetic-propagation characteristics in frequency bands already in use may be overcome. These trends include the use of meteor

trails (getting reflections of radio waves from the highly ionized tail following a meteor particle entering the upper atmosphere), the use of the Moon's surface as a passive reflector of radio waves, retransmission of signals by means of artificial Earth's satellites. The latter is of special importance for effecting reliable communication and radio- and telecasting. It may be supposed that such methods will enable the development of a communication system combining the advantages of the different frequency bands.

1.3. FUNDAMENTAL RADIO PROCESSES

From the foregoing it is clear that any radio signal transmitted through a communication channel undergoes considerable changes. Some of these processes are typical of the majority of communication systems, regardless of their design and the character of the message being transmitted. These fundamental processes and their

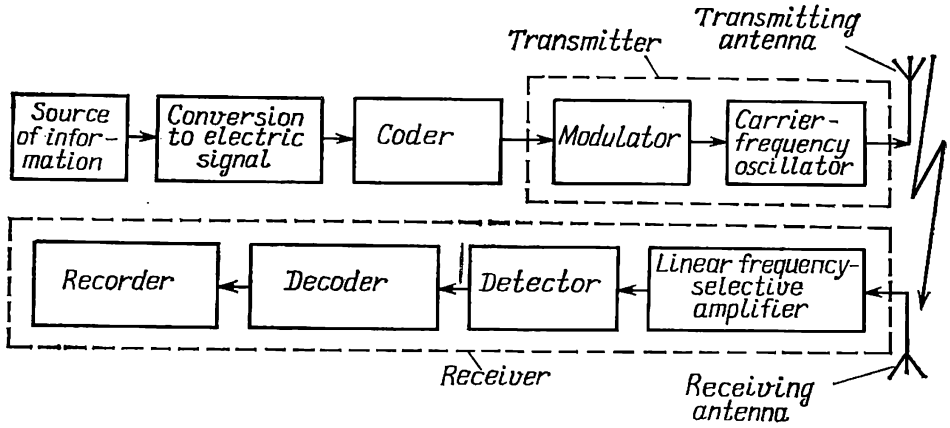


Fig. 1.1. Radio communication channel

basic features are briefly outlined in the general block diagram of a radio channel given in Fig. 1.1.

Conversion of the initial message into a coded electric signal. When transmitting speech or music, the sound waves are converted into electrical oscillations by means of a microphone, while optical images for television transmission are transformed into video signals by means of a camera tube (e.g., superorthicon). If a written message is to be transmitted, it is first coded, i.e., each letter of the text is replaced by a combination of standard symbols (such as dots, dashes and pauses of the Morse code) and these are then converted into standard electrical signals (e.g., pulses of different duration or polarity).

It should be noted that the diagram shown in Fig. 1.1 illustrates a case where the information is inserted at the "beginning" of the communication system, i.e., directly in the transmitter. This is not so in the case of a radar channel, where the target data (range, height, velocity, etc.) is received as radio waves reflected from the target in free space.

Generation of radio-frequency oscillations. A radio-frequency oscillator is a source of carrier waves. Depending on the purpose and type of radio communication channel, the oscillator may have an output from a few thousands of a watt to millions of watts. Obviously, such oscillators differ in design and overall dimensions very broadly—from a small element to a huge engineering structure.

The basic characteristics of a radio-frequency oscillator are: frequency range and capability of being quickly transferred from one operating frequency to another, power output, and efficiency. The stability of oscillation frequency is of special importance. In this respect, the position of radio engineering is unique. The electromagnetic-propagation characteristics and the wide frequency spectrum of signals necessitate the use of very high carrier frequencies. On the other hand, the fact that the signals have to be processed in the presence of interfering signals and the need to suppress interstation interference require that the absolute value of the frequency fluctuations should be minimized. Hence the extremely stringent requirements for the *relative* frequency stability.

Control of radio-frequency oscillattions (*modulation*). The process of modulation consists in changing one or several parameters of a radio-frequency wave in accordance with the information to be transmitted. As a rule, the frequency of the modulating signal is low in comparison with the carrier frequency of the oscillator. Modulation can be effected by various methods, usually based on changing the potential applied to the electrodes of electron devices inserted into the circuit of the radio transmitter. The principal characteristic of the modulation process is the degree of conformity between the change in a parameter of the high-frequency oscillation and the modulating signal.

Reception and amplification of weak signals. The antenna of a radio receiver "catches" just a minute fraction of the energy radiated by the transmitting antenna system. Depending on the distance between the transmitter and receiver, directivity patterns of the antennas, and propagation conditions, the input power of the receiver is about 10^{-10} to 10^{-14} W, while for reliable reproduction of the signal the output power must be of the order of several milliwatts to several watts or higher. Therefore, the receiver must ensure a gain of 10^7 to 10^{14} in power or 10^4 to 10^7 in voltage.

Modern receivers provide steady reception of radio signals at an input voltage of about 1 μV. Such a high sensitivity of up-to-date

receivers has become possible due to the latest achievements in radioelectronics. Of fundamental importance are also special methods of designing the receiver circuits to provide a high gain and ensure steady operation of all the units in the system. These methods allow one to convert (lower) the oscillation frequency in the receiver channel without breaking the structure of the signal being received (in Fig. 1.1 the frequency conversion process is not shown).

In addition to the receiving systems, the frequency conversion process is widely used in various radio engineering and measuring devices. The problem of amplifying weak signals in a receiver is intimately associated with the problem of detecting them in the midst of random noise. Therefore, one of the basic parameters of any receiver is its selectivity or its ability to differentiate the desired signal from interfering signals broadcasted on other frequencies.

The frequency selectivity is accomplished with the aid of resonance circuits.

Separation of the message from a high-frequency oscillation (detection and decoding). The detection or demodulation is the reverse of the modulation. This process consists in obtaining a voltage (current) that varies in time according to the variations of some parameter (amplitude, frequency or phase) of the modulated wave. In other words, detection results in the restoration of the original message. The detector is usually inserted into an output stage of the receiver. Consequently, it processes a modulated oscillation already amplified in the preceding stages. The main requirement for the detector is the exact reproduction of the form of the original signal.

After detection, the decoding of the signal is effected, which is the reverse of the coding. In many radio channels coding and decoding are not used.

Besides the above processes associated in some way or other with the conversion of frequency spectra, radio engineering often employs amplification without frequency conversion, effected by means of amplifiers. These include:

— low-frequency control-signal amplifiers inserted before the modulator of the transmitter and also at the output of the receiver;

— short-pulse amplifiers used in television, radar, and pulsed radio communication systems;

— high-power, high-frequency amplifiers employed in radio transmitters;

— radio-frequency weak-signal amplifiers employed in radio receivers and measuring instruments.

In addition to the above-mentioned processes, which, as it has been already stated, are typical of all radio engineering systems, many other processes, such as frequency multiplication and division, short-pulse generation, and various types of pulse modulation, are used in special cases.

1.4. ANALOG, DISCRETE AND DIGITAL SIGNALS AND CIRCUITS

The block diagram of a communication channel shown in Fig. 1.1 gives no information on the type of signal used for transmitting a message and on the structure of the individual units of the channel.

The signals used in modern radioelectronics may be divided into the following classes:

— continuous signals of arbitrary magnitude (Fig. 1.2*a*);

— time-quantized (sampled) signals of arbitrary magnitude (Fig. 1.2*b*);

— level-quantized (or simply quantized) continuous signals (Fig. 1.2*c*);

— quantized and sampled signals (Fig. 1.2*d*).

Signals of the first class (Fig. 1.2*a*) are sometimes called *analog signals* (since they can be regarded as electrical models of physical

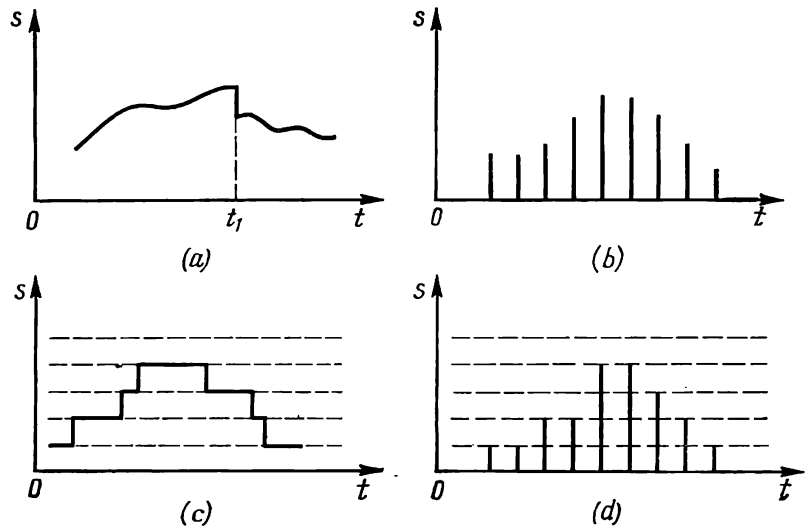


Fig. 1.2. Signals: (*a*) continuous signals of arbitrary magnitude; (*b*) sampled signals of arbitrary magnitude; (*c*) quantized continuous signals; (*d*) quantized and sampled signals

quantities) or *continuous signals* (since they are specified along the time axis on a nonenumerable set of points). In this case they can have any magnitude within a definite range on the ordinate axis. Since these signals can have discontinuities, as shown in Fig. 1.2*a*, they would better be specified as "*continual signals*" to avoid incorrect description.

Figure 1.2*b* shows a signal specified at discrete moments of time t (on an enumerable set of points), whose magnitude at these points can be anything within a definite interval on the ordinate axis

(as in Fig. 1.2*a*). Thus, the term "discrete" characterizes not the signal itself, but the method of its specification on the time axis.

From Fig. 1.2*c* it is clear that the signal here is specified along the whole time axis, but it can have only discrete magnitudes. In this case we have a *level-quantized signal.* Later in the text we shall use the term "*discrete*" only for time-quantized or sampled signals, while level-quantized signals will simply be referred to as *quantized.*

Quantization is used for presenting signals in a digital form through digital coding, since the levels (signal magnitudes) can be numbered

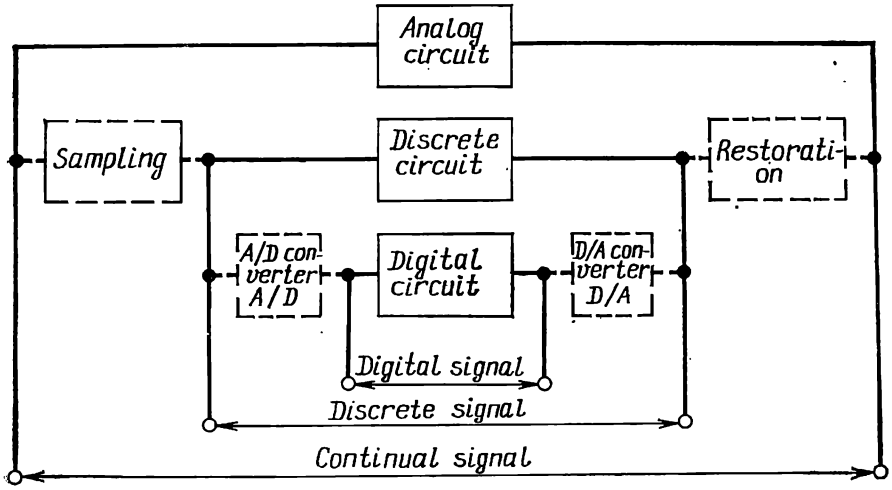


Fig. 1.3. Types of signals and the corresponding circuits

by digits with a finite number of places. Therefore, later in the text, a sampled and quantized signal (Fig. 1.2*d*) will be frequently referred to as *digital signal.*

Thus, we may distinguish between continual (Fig. 1.2*a*), discrete (Fig. 1.2*b*), quantized (Fig. 1.2*c*), and digital signals.

Each of these signals can be correlated with an analog, a discrete, or a digital circuit. The relation between the type of signal and the type of circuit is shown in the block diagram of Fig. 1.3*.

When processing a continual signal by means of an analog circuit, no additional signal conversion is necessary. On the other hand, the processing of a continual signal by means of discrete circuit requires two conversions: (a) signal sampling at the input of the discrete circuit and (b) inversion, i.e., restoration of the continual structure of the signal at the output of the discrete circuit. Finally,

* The above-mentioned correction of terminology and classification of signals and circuits according to the diagram shown in Fig. 1.3 was proposed by A. M. Trakhtman.

the digital processing of a continual signal requires two more stages of conversion: (a) analog-to-digital conversion, i.e., quantization and digital coding, at the input of the digital circuit and (b) digital-to-analog conversion, i.e., decoding, at the input of the digital circuit.

It should be noted that at present the digital processing of signals is gaining ever wider application not only because of the high accuracy and versatility of this method, but also because of the possibilities offered by modern developments in microelectronics.

The discrete and digital signals are discussed in Chapter 13.

1.5. METHODS OF ANALYSIS OF RADIO CIRCUITS

Any radio engineering system is composed of a large number of linear and nonlinear elements and circuits. Linear circuits are subclassified into *circuits with constant parameters* and *circuits with variable parameters.*

Furthermore, each of the above-mentioned classes is subdivided into *circuits with lumped parameters* and *circuits with distributed parameters.* The former include circuits composed of inductances, capacitances and resistances, while the latter include circuits comprising lines, waveguides and energy radiating devices.

This course deals mainly with circuits having lumped parameters. The basic characteristics of these circuits will be readily understood upon recalling the properties of the differential equations describing them. Having in mind the circuits with lumped parameters, let us write the following three equations:

$$a_0 \frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + a_2 \frac{d^{n-2} y}{dt^{n-2}} + \ldots + a_{n-1} \frac{dy}{dt} + a_n y = f(t) \qquad (1.1)$$

$$a_0 \frac{d^n y}{dt^n} + a_1(t) \frac{d^{n-1} y}{dt^{n-1}} + a_2 \frac{d^{n-2} y}{dt^{n-2}} + \ldots + a_{n-1} \frac{dy}{dt} + a_n y = f(t) \qquad (1.2)$$

$$a_0 \frac{d^n y}{dt^n} + a_1(y) \frac{d^{n-1} y}{dt^{n-1}} + a_2 \frac{d^{n-2} y}{dt^{n-2}} + \ldots + a_{n-1} \frac{dy}{dt} + a_n y = f(t) \qquad (1.3)$$

Equation (1.1) is a linear equation with constant coefficients, a_0, a_1, a_2, . . ., a_n, characterizing a linear circuit with constant parameters. Equation (1.2), in which at least one coefficient [in this case, $a_1(t)$] is a function of time (but not of y) is a linear equation with a variable coefficient (or coefficients) and describes a linear circuit with variable parameters. Finally, equation (1.3), in which one or several coefficients [namely $a_1(y)$] are functions of y, is a nonlinear differential equation describing a nonlinear circuit.

First, let us consider the properties of a linear equation with constant coefficients. For better understanding, let the general equation (1.1) be replaced by a simple second-order equation such

as that pertaining to a series circuit L, C, r to which an electromotive force $e(t)$ is applied.

The current $i(t)$ flowing through the circuit can be expressed by the following integro-differential equation

$$L\frac{di}{dt}+ri+\frac{1}{C}\int idt=e(t) \tag{1.4}$$

Differentiating this equation with respect to t and taking $e'(t) = f(t)$, we come to the equation (1.1).

Equation (1.4) is linear if the factors L, r and $1/C$ *do not depend on the current* i or, just the same, *upon the magnitude of the external force* $e(t)$. When this condition is satisfied, the voltages across each element of the circuit are linearly related to the current. In fact, denoting these voltages as v_r, v_L, and v_C respectively, we may write

$$v_r=ri, \quad v_L=L\frac{di}{dt}, \quad v_C=\frac{1}{C}\int idt \tag{1.5}$$

Since the differentiation and integration are linear operations, we can state that v_L and v_C are linearly related to the current i whatever the law governing the variation of the latter in time. With respect to v_r this statement is still more obvious.

One of the manifestations of circuit linearity is the independence of the relation between the input and output voltages (currents) of the input voltage (current).

For example, when the current varies as $i = I \sin \omega t$, we obtain

$$v_r=rI\sin\omega t; \quad v_L=\omega LI\cos\omega t; \quad v_C=-\frac{1}{\omega C}I\cos\omega t \tag{1.6}$$

A change in the current amplitude I results in the same change in the voltage amplitudes across the elements r, L, and C. This property of linear elements may be interpreted as resulting from their linear current-voltage characteristics. The current-voltage characteristic for the element r is given in Fig. 1.4, where both instantaneous (v_r and i) and amplitude values can be plotted along the coordinate axes, while the characteristic for the elements L and C is given in Fig. 1.5, where the amplitudes V_L, I or V_C, I are plotted. It should be noted that current-voltage characteristics express the relation between the amplitudes V_L, V_C and I without regard to their dependence on frequency. As will be recalled, the functions $i(t)$ and $v_r(t)$ for the element r can differ only in the constant coefficient $1/r$ which is numerically equal to the slope of the current-voltage characteristic (see Fig. 1.4):

$$\tan\alpha = Ni/v_r = N/r$$

The slopes of the current-voltage characteristics $I\ (V_L)$ and $I\ (V_C)$ plotted for a fixed frequency ω are:
for inductance

$$\tan \alpha = NI/V_L = N/\omega L$$

for capacitance

$$\tan \alpha = NI/V_C = N/\omega C$$

In these expressions the coefficient N having the dimension of resistance depends on the scales along the abscissa and ordinate

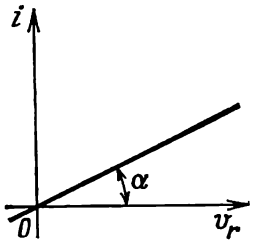


Fig. 1.4. Current-voltage characteristic of a linear resistor

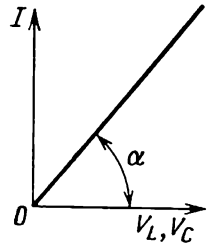


Fig. 1.5. Current-voltage (amplitude) characteristic of a linear element (L or C)

axes. If voltages, currents and resistances are expressed respectively in volts, amperes and ohms, $N = 1$ ohm.

Another important property of linear circuits that also follows from the linearity of the differential equation describing their behaviour (current, voltage) is the fact that they obey *the principle of independence or superposition.* This principle may be defined as follows: *if several external forces act on a linear circuit, its behaviour (current, voltage) can be determined by superposing the solutions found separately for each force.* Another definition reads: *in a linear circuit the sum of the effects of various external forces is equal to the effect of the sum of these forces.* In this case it is assumed that the circuit has no initial store of energy.

The superposition principle serves as a basis for the spectrum and operator methods of analysis of transient processes in linear circuits, as well as for Duhamel's superposition integral. By applying the superposition principle, any complex signal transmitted through a linear circuit can be resolved into simple signals, say sinusoidal, which are more convenient to analyze.

Now let us consider another fundamental characteristic of a linear circuit, that stems directly from the theory of integration of linear differential equations with constant coefficients. Using a Fourier series or integral for resolving $e\ (t)$ on the right-hand side of equation (1.4) into simple harmonic components acting when $-\infty <$

$< t < \infty$, we obtain the solution of equation (1.4) for each component with a frequency ω_n in the form of a harmonic oscillation of the same frequency:

$$i_n(t) = I_n \cos(\omega_n t + \theta_n)$$

where I_n and θ_n are constant amplitude and phase.

From this it follows that *no complex action upon a linear circuit with constant parameters produces new frequencies in this circuit.* This means that no conversion of signals is accompanied by the appearance of new frequencies (i.e., those not present in the spectrum of the input signal) and that no frequency conversion can in principle be performed with the aid of a linear circuit with constant parameters. Linear circuits with constant parameters, which are used in radio engineering, help to solve problems not connected with the transformation of spectrum, such as linear amplification of signals and filtering (according to the required frequency).

Now let us study the characteristics of linear circuits with variable parameters in view of the general equation (1.2).

As in the preceding case, we can rely on the superposition principle. This means that the right-hand side of equation (1.2), i.e., the external force $f(t)$, can be resolved into harmonic components acting when $-\infty < t < \infty$. The solution of equation (1.2) is then obtained as a sum of independent partial solutions corresponding to each of the components of the right-hand side of this equation. (As before, it is assumed that no energy has been stored in the circuit.) However, in contrast to the preceding case, these partial solutions are not harmonic, but more complex functions. In other words, even a simplest harmonic action on a linear circuit with variable parameters produces a complex oscillation having a *spectrum of frequencies.* This can be illustrated by the following simple example. Let a harmonic voltage

$$e(t) = E_0 \cos \omega t$$

be applied to a resistor whose resistance changes in time as

$$R(t) = R_0/(1 + M \cos \Omega t)$$

The current flowing through the resistor is

$$i(t) = e(t)/R(t) = (E_0/R_0)(1 + M \cos \Omega t) \cos \omega t$$
$$= (E_0/R_0)[\cos \omega t + (M/2) \cos(\omega + \Omega) t + (M/2) \cos(\omega - \Omega) t] \quad (1.7)$$

As is seen from this equation, the current includes components with frequencies $\omega \pm \Omega$ which are not present in $e(t)$. Even this simple model shows that the spectrum of the input signal can be converted by making the resistance vary in time.

Similar results can be obtained, though through more complicated computations, for a circuit with variable parameters, containing reactive elements: inductances and capacitances. This problem is dealt with in Chapter 10. Here, it should be only noted that a linear circuit with variable parameters converts the frequency spectrum of the acting signal and, therefore, can be used for signal conversion accompanied by the transformation of spectrum. From further discussion it will be clear that by making the inductance or capacitance of an oscillatory circuit vary periodically in time one can (under certain conditions) "pump" energy from an auxiliary source ("parametric amplifiers" and "parametric oscillators". Chapter 10).

The theory of differential equations with variable coefficients is much more complicated than that of equations with constant coefficients. Even in the case of a harmonic right-hand side, the solution of second (or higher) order equations can be found only for a few particular cases. Hence, it is clear that, although the superposition principle is applicable to linear circuits with variable parameters, the spectrum analysis cannot be applied to the transmission of signals through such circuits in all cases. This problem is discussed in detail in Chapter 10.

Finally, let us discuss the general properties of nonlinear circuits. From the theory of nonlinear differential equations we know that the superposition method is not suitable for solving such equations. This means that if an equation similar to (1.3) (for example at $n=1$) and with $f_1(t)$ on the right-hand side

$$a_1(y)\frac{dy}{dt}+a_0y=f_1(t)$$

has a solution in the form of a function $y_1(t)$, and another similar equation with $f_2(t)$ on the right-hand side

$$a_1(y)\frac{dy}{dt}+a_0y=f_2(t)$$

has a solution $y_2(t)$, then the equation

$$a_1(y)\frac{dy}{dt}+a_0y=f_1(t)+f_2(t)$$

has a solution in the form of a function $y_3(t)$ which is not equal to the sum of $y_1(t)$ and $y_2(t)$, i.e.,

$$y_3(t)\neq y_1(t)+y_2(t)$$

Thus, it is clear that the superposition principle is inapplicable to nonlinear elements and circuits. This property of nonlinear circuits is intimately connected with the curvature of the current-voltage (or similar) characteristics of nonlinear elements. Figure 1.6

illustrates a typical diode characteristic which, in contrast to the current-voltage characteristic of a linear resistor (Fig. 1.4), manifests no direct proportionality between current and voltage. If a current i_1 corresponds to a voltage v_1 and a current i_2 corresponds to a voltage v_2, the total voltage $v_3 = v_1 + v_2$ produces a current i_3 that is not equal to the sum $i_1 + i_2$ (see Fig. 1.6).

It is obvious from this simple example that when analyzing the action of a compound signal on a nonlinear circuit, this signal cannot be resolved into simple components; one has to find the response of the circuit to the resultant signal.

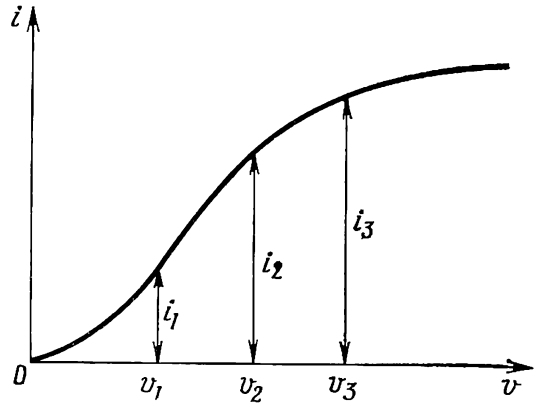


Fig. 1.6. Current-voltage characteristic of a nonlinear element (diode)

The inapplicability of the superposition principle to nonlinear circuits makes unfit the spectrum analysis, as well as other analysis methods based on the resolution of a complex signal into its components.

Another important characteristic of a nonlinear circuit is the *transformation of signal spectrum.*

When a simple harmonic signal acts on a nonlinear circuit, there appear, in addition to the fundamental-frequency component, some harmonics, i.e., components having frequencies that are integral multiples of the fundamental frequency (in some cases, a d-c current or voltage component also appears).

It will be shown later in the text that in the case of a compound signal, in a nonlinear circuit there also develop combination-frequency components resulting from the interaction of some frequency components of the signal.

From the viewpoint of signal spectrum conversion, we should emphasize the fundamental difference between *linear parametric circuits* and *nonlinear circuits.* In a nonlinear circuit the output spectrum structure depends not only on the form of the input signal but on its amplitude as well, while in a linear parametric circuit the spectrum structure is independent of the signal amplitude.

Free oscillations in nonlinear circuits are of special interest for radio engineers. Such oscillations are commonly referred to as *self-excited oscillations*, because they develop and may stably exist without external excitation. A d-c source compensates for the energy consumption in such a circuit.

The fundamental radio processes — generation, modulation, detection, and frequency conversion — are accompanied by the transformation of frequency spectrum. Therefore, these processes can be effected through the use of either nonlinear circuits or linear circuits with variable parameters. In some cases, both nonlinear and linear parametric circuits are employed simultaneously. It should be also stressed that nonlinear elements generally operate in conjunction with linear circuits which separate the useful components of the transformed spectrum. In this connection, the classification of circuits as linear, nonlinear and parametric is rather conventional. When analyzing actual radio circuits including nonlinear elements, one has to use various mathematical methods, both linear and nonlinear, to describe the behaviour of the individual units of one and the same device.

It should also be noted that even in the case of linear circuits, different methods are employed to analyze the circuits with lumped parameters and those with distributed parameters. The use of some or other radio circuit is decided by the operating frequency band. Hence it is clear that no detailed classification of radio circuits can be made without due regard for the frequency band used.

1.6. NOISE IMMUNITY OF COMMUNICATION CHANNELS

As has been stated above, radio engineering deals with the transmission and reception of information. The set of devices used to convey information from its source to the point of reception (and also the medium therebetween) forms a *communication channel.* The communication channel must ensure that the messages be transmitted as fully as possible, i.e., with minimum losses. The loss of information can be caused by the distortion of the signals due to the *imperfection of individual components of the channel* and *noise* or *interference.*

Noise arises in all the elements of a communication channel, both in the medium used for propagating the signal from the transmitter to the receiver and in the equipment used for the required transformations of the signal. In the former case it is called *external noise* and in the latter, *internal noise.*

The external noise is due either to various atmospheric phenomena (lightning discharges, electrification of particles on account of friction, etc.) and extraterrestrial sources (solar and stellar radio emission) or to industrial interference (sparking in current collectors, electric welding, commutation of power units and networks, ignition systems of internal combustion engines, etc.). Interference signals are also produced during the operation of medical equipment, such as X-ray apparatus and physiotherapeutic devices. Radio stations operating at adjacent frequencies also interfere with the

transmitted signal. The interference may be deliberate, as the noise produced for jamming radio broadcasts.

The internal (natural) noise is caused by the thermal motion of discrete charged particles in the elements of electric circuits, the shot effect in electron devices, and other phenomena taking place during the operation of radio engineering systems. The internal noise is particularly detrimental to the reception of weak signals which have to be considerably amplified. The interfering signals are amplified together with the useful signal, and the noise might have an intensity commensurable with that of the signal, so that the latter is partially or completely blurred.

The most effective means of noise suppression is cancelling out or attenuating the interfering signals at their origin. For this purpose (as in the case of industrial noise sources) one must check the condition of electric contacts and adjust them or utilize electrostatic shields, spark arresters, filters, etc. The elimination of radio interference can be achieved by correctly distributing the working frequencies, which is regulated by special international agreements, improving the quality of transmission through the suppression of undesirable (parasitic) radiation, providing highly stable carrier frequencies, using directional antennas, etc. All these measures help to solve the problem of "overcrowded bands". For the same purpose a frequency band should be selected, which has the lowest noise level.

The problem of reducing natural noise is very complicated, but the noise level can be significantly lowered by using amplifiers which are deeply cooled (e.g., down to the temperature of liquid helium), so that the rate of thermal (random) motion of particles is considerably reduced. None the less, in spite of all these measures, the noise background cannot be eliminated completely. A certain amount of circuit noise, cosmic noise, static, etc. is always present. The distortion of the useful signal under the effect of noise is dangerous in that the unique relation between the transmitted and received signals is disturbed on account of the random character of the interference and becomes merely more or less probable. During reception, the distortion of the signal may result in replacement of one message (i.e., the true transmitted information) by another, probable message, which will in this case contain false information. The operator at the receiving end is therefore not sure of the authenticity of the received message, and the reception thus becomes *unreliable*. Therefore, the *problem of noise immunity* still remains a basic problem of radio engineering. A communication system must be designed so as to be capable of withstanding in the best way the harmful effects of interference.

The problem of noise immunity includes a wide variety of associated problems: generation of high-power oscillations, selection of

frequency bands which ensure favourable conditions for propagating electromagnetic waves, use of directional antennas, study of new types of radio signals and new ways of detecting them in the midst of random noise, etc.

Since any kind of interference is, as a rule, a random process, the problem of noise immunity cannot be successfully solved without applying the methods of the theory of probability and the theory of random functions. These methods have become particularly important for radio engineering after the development of the general theory of communication which in essence is a *statistical theory*.

1.7. OBJECTIVES OF THE PRESENT COURSE

The principal objective of this course is to explain the physical processes occurring in radio engineering systems and to provide a mathematical description of these processes.

In accordance with this aim, the course includes:

(1) analysis of deterministic oscillations and random processes—signals and noise;

(2) analysis of radio circuits—linear, nonlinear and parametric;

(3) analysis of the propagation of signals through radio circuits;

(4) theory of the main radio engineering transformations—generation of oscillations, modulation, detection, frequency conversion, frequency multiplication;

(5) main propositions of the theory of synthesis of linear radio circuits;

(6) introduction to the analysis and synthesis of discrete and digital circuits.

The material in this book is arranged in the following order:

Chapters 2 and 3 deal with the characteristics of the deterministic signals—control and modulated radio-frequency oscillations, while Chapter 4 discusses the basic characteristics of random signals. Chapter 5 considers linear circuits comprising active elements, and also circuits with feedback loops and their stability. Chapters 6 and 7 deal with the transmission of signals, both deterministic and random, through linear radio circuits with constant parameters. Chapters 8 and 9 cover nonlinear circuits and self-oscillating systems, while Chapter 10 describes parametric circuits. Chapter 11 discusses the action of random processes on nonlinear and parametric circuits. Chapter 12 deals with the matched filtering of signals contaminated by noise, while Chapter 13 considers discrete and digital processing of signals. Chapter 14 gives brief information on the widely used methods of approximation of characteristics and processes.

The problems associated with the synthesis of analog and digital filters are discussed in Chapter 15.

Chapter 2

SIGNALS

2.1. GENERAL

Radio engineering deals with electric signals that are modified by some method of coding to contain or represent the messages to be conveyed. Such a signal, therefore, can be regarded as an information-bearing physical (electrical) process. The amount of information that can be transmitted by an electric signal depends on its principal parameters: duration, frequency band, power, and so on. The noise level in the communication channel is of prime importance: the lower the noise level, the greater the amount of information that can be transmitted by an electric signal of specified power. Before discussing the informative value of a signal, let us consider its basic characteristics. In this connection, it is expedient to discuss separately the *deterministic* and *random* signals.

A signal whose instantaneous value can be predicted for any moment of time with a unit probability is called a *deterministic* signal. For example, pulses or trains of pulses of known shape, magnitude and position in time, or a continuous signal with given amplitude and phase relations within its spectrum may be considered deterministic signals. Deterministic signals may be *periodic* and *nonperiodic*.

A *periodic* signal is one satisfying the condition $s(t) = s(t + kT)$, where period T has a finite length and k is any integer.

A simple periodic deterministic signal is a harmonic oscillation (current, voltage, charge, field strength) obeying the law

$$s(t) = A \cos(2\pi t/T + \theta) = A \cos(\omega t + \theta), \quad -\infty < t < \infty \tag{2.1}$$

where A, T, ω and θ are the constant amplitude, period, angular frequency, and epoch angle, i.e., the initial phase, of the oscillation.

A purely harmonic oscillation is referred to as a *monochromatic wave*. This term, adopted from optics, indicates that the spectrum of the harmonic oscillation consists of a single spectral line. Spectra of real signals having a beginning and an end are inevitably blurred. So, pure monochromatic waves do not exist in nature. Therefore, under a harmonic and monochromatic signal we shall conventionally imply a signal determined by a function coinciding with expres-

sion (2.1) within a finite interval wide enough to neglect the influence of the "wings".

It is generally known that any compound periodic signal can be presented as a sum of harmonic waves with frequencies that are multiples of the fundamental frequency $\omega = 2\pi/T$. The main characteristic of a compound periodic signal is its spectral function containing the information about the amplitudes and phases of individual harmonics.

Any deterministic signal for which the condition $s(t) = s(t + kT)$ is not satisfied is called *nonperiodic deterministic* signal.

A nonperiodic signal has, as a rule, a limited duration. The above-mentioned pulses, pulse trains, "fractions" of harmonic waves may serve as examples of such signals. Nonperiodic signals are of particular interest since they are most widely used in radio engineering practice.

The spectral function of a nonperiodic (as well as a periodic) signal is its main characteristic; the spectral structure of a nonperiodic signal has some specific features that will be discussed in detail later in the text.

Random signals are ones whose magnitudes are not known beforehand and can only be predicted with a probability less than unity. As examples of such signals we can name a voltage corresponding to voice or music, or sequences of telegraph signals conveying a non-repeated text. Series of radio pulses at the input of a radar receiver are also random signals, since their high-frequency carriers fluctuate in amplitude and phase depending on the conditions of propagation, position of the target, etc. Many other examples of random signals can be given. In fact, any information-bearing signal should be regarded as a random signal. The "completely known" deterministic signals contain no information. Later in the text such signals will often be called "oscillations".

A statistical approach is used for specifying and analyzing random signals. Among the main characteristics of random signals are the following: (a) the law of probability distribution and (b) the spectral distribution of signal power.

The first characteristic can be used for determining the relative interval of time the signal falls into a definite interval of levels, the ratio of the maximum signal value to the rms value (peak factor), and some other important parameters of the signal. The second characteristic presents only the frequency distribution of the *average signal power*. The spectral characteristic of a random process does not give any detailed information on the individual components of the spectrum, e.g., on their amplitudes and phases.

Besides the useful random signals, the theory and practice of radio engineering have to deal with sporadic interfering signals, or "noise". As mentioned above, the noise level is a principal factor

that limits the amount of information that can be conveyed by a given signal. Therefore, the study of random signals is inseparable from the study of noise. These problems are discussed in Chapter 4.

2.2. EXPANSION OF AN ARBITRARY SIGNAL IN A GIVEN SET OF FUNCTIONS

The expansion of a given function in various orthogonal sets of functions plays an important role in the theory and practice of forming and processing radio signals. In this connection, let us recall the fundamental relations defining the properties of orthogonal sets.

An infinite set of real functions

$$\varphi_0(x),\ \varphi_1(x),\ \varphi_2(x),\ \ldots,\ \varphi_n(x),\ \ldots \tag{2.2}$$

is called orthogonal in an interval $[a, b]$ if

$$\int_a^b \varphi_n(x)\,\varphi_m(x)\,dx = 0 \quad (n \neq m) \tag{2.3}$$

In this case it is assumed that

$$\int_a^b \varphi_n^2(x)\,dx \neq 0 \tag{2.4}$$

i.e., none of the functions of the set (2.2) is identically equal to zero.

Condition (2.3) expresses the pairwise orthogonality of the functions belonging to the set (2.2).

The quantity

$$\| \varphi_n \| = \sqrt{\int_a^b \varphi_n^2(x)\,dx} \tag{2.5}$$

is called the norm of the function $\varphi_n(x)$.

The function $\varphi_n(x)$ satisfying the condition

$$\| \varphi_n \|^2 = \int_a^b \varphi_n^2(x)\,dx = 1 \tag{2.6}$$

is called *normalized* function, while a set of normalized functions $\varphi_1(x)$, $\varphi_2(x)$, . . ., in which any two different functions are mutually orthogonal is called *orthonormal* set.

It is known from mathematics that if functions $\varphi_n(x)$ are continuous, an arbitrary piecewise-continuous function $f(x)$ satisfying the condition

$$\int |f(x)|^2 dx < \infty \tag{2.7}$$

may be presented as a sum of a series:

$$f(x) = c_0\varphi_0(x) + c_1\varphi_1(x) + \ldots + c_n\varphi_n(x) + \ldots \tag{2.8}$$

The integral in expression (2.7) is calculated in the domain of definition of the function $f(x)$.

Let us multiply both parts of equation (2.8) by $\varphi_n(x)$ and integrate between a and b. All terms of the form $\int_a^b c_m\varphi_m(x)\,\varphi_n(x)\,dx$ are zero if $m \neq n$, because of the orthogonality of the functions $\varphi_m(x)$ and $\varphi_n(x)$.

The only term remaining on the right-hand side of the equation is

$$\int_a^b c_n\varphi_n(x)\,\varphi_n(x)\,dx = c_n\int_a^b \varphi_n^2(x)\,dx = c_n \|\varphi_n\|^2$$

so we may write

$$\int_a^b f(x)\,\varphi_n(x)\,dx = c_n \|\varphi_n\|^2$$

from which follows an important expression

$$c_n = \frac{1}{\|\varphi_n\|^2}\int_a^b f(x)\,\varphi_n(x)\,dx \tag{2.9}$$

The series (2.8) with the coefficients c_n defined by equation (2.9) is called *generalized Fourier series* in a given set of $\varphi_n(x)$. The generalized Fourier series has the following important property: for a given set of functions $\varphi_n(x)$ and with a certain definite number of terms in the series (2.8), it gives the best approximation (i.e., with a minimum mean-square error) of the given function $f(x)$. This means that the mean-square error

$$M = \int_a^b \left[f(x) - \sum_{n=0}^{N} a_n\varphi_n(x)\right]^2 dx$$

is minimal when $a_n = c_n$. Indeed, substituting $a_n = c_n + b_n$ into the above expression and using equations (2.3), (2.5), and (2.9), we obtain

$$M = \int_a^b f^2(x)\,dx - \sum_{n=0}^{N} c_n^2 \| \varphi_n \|^2 + \sum_{n=0}^{N} b_n^2 \| \varphi_n \|^2$$

From this it follows that M is minimal when $b_n = 0$, i.e., when $a_n = c_n$. Thus,

$$M_{\min} = \int_a^b f^2(x)\,dx - \sum_{n=0}^{N} c_n^2 \| \varphi_n \|^2 \tag{2.10}$$

Since the quantity

$$\int_a^b f^2(x)\,dx = \| f \|^2$$

is the square of the norm of the function $f(x)$ and $M_{\min} \geqslant 0$, expression (2.10) yields the following inequality

$$\sum_{n=0}^{N} c_n^2 \| \varphi_n \|^2 \leqslant \| f \|^2 \tag{2.11}$$

This principal inequality, called *Bessel's inequality*, is true for any orthogonal set.

An orthogonal set is *complete* if the mean-square error M can be made as small as desired by increasing the number of terms in the series.

The condition of completeness may be written as

$$\sum_{n=0}^{\infty} c_n^2 \| \varphi_n \|^2 = \| f \|^2 \tag{2.12}$$

If this condition is satisfied, it is safe to assume that the series (2.8) *converges in mean*, i.e.,

$$\lim_{N \to \infty} \int \left[f(x) - \sum_{n=0}^{N} c_n \varphi_n(x) \right]^2 dx = 0 \tag{2.13}$$

It does not mean, however, that $\sum_{n=0}^{\infty} c_n \varphi_n(x)$ converges to $f(x)$, i.e.,

$$\max \left| f(x) - \sum_{n=0}^{\infty} c_n \varphi_n(x) \right| = 0$$

for any values of x. In Section 2.4 an example is given showing that the series $\sum_{n=0}^{\infty} c_n \varphi_n(x)$ may differ from $f(x)$ at some points on the x-axis, although equality (2.13) holds.

For a set of complex functions the above definitions can be generalized in the following manner:

The condition of orthogonality

$$\int_a^b \varphi_n(x)\, \varphi_m^*(x)\, dx = 0 \quad (n \neq m) \tag{2.3'}$$

The square of the norm of the function $\varphi_n(x)$

$$\| \varphi_n \|^2 = \int_a^b \varphi_n(x)\, \varphi_n^*(x)\, dx = \int_a^b |\varphi_n(x)|^2\, dx \tag{2.5'}$$

Fourier coefficients

$$c_n = \frac{1}{\| \varphi_n \|^2} \int_a^b f(x)\, \varphi_n^*(x)\, dx \tag{2.9'}$$

In these expressions the asterisk denotes the complex conjugate.

Later in the text, when applied to signals $s(t)$ which are functions of time, expression (2.8) will be written as

$$s(t) = \sum_{n=0}^{\infty} c_n \varphi_n(t) \tag{2.14}$$

In this case relation (2.12) has the meaning of energy. In fact, after replacing $f(x)$ by $s(t)$, the quantity $\| f \|^2$ may be written as

$$\| s \|^2 = \int_{t_1}^{t_2} s^2(t)\, dt = E \tag{2.15}$$

If $s(t)$ stands for an electric signal (voltage, current), then E is the energy of the signal in the interval $t_2 - t_1$ (provided that the energy is dissipated in a resistor of 1 ohm).

Thus, in agreement with formula (2.12), the signal energy is

$$E = \sum_{n=0}^{\infty} c_n^2 \| \varphi_n \|^2 \tag{2.16}$$

whereas the use of an orthonormal set of functions $\varphi_n(t)$ yields

$$E = \sum_{n=0}^{\infty} c_n^2 \tag{2.16'}$$

In this case it should be kept in mind that the time interval $t_2 - t_1$ for which the energy E is determined is the interval of orthogonality for the set of functions $\varphi_n(t)$.

It is evident that the average power of the signal over the time interval $t_2 - t_1$ is

$$\overline{s^2(t)} = \frac{E}{t_2 - t_1} = \frac{1}{t_2 - t_1} \sum_{n=0}^{\infty} c_n^2 \| \varphi_n \|^2 \qquad (2.17)$$

The choice of the most rational orthogonal set of functions depends upon the purpose of expanding a compound function in a series. Among the numerous problems requiring expansion of a complex function, most important are as follows:

(a) exact expansion into simple orthogonal functions;

(b) approximation of signals, processes, or characteristics when the number of terms in the series must be reduced to a minimum (with a specified permissible error).

The first problem is generally solved through the use of the orthogonal set of basic trigonometric functions—sines and cosines. This is preferable for many reasons. First, a harmonic oscillation is the only function of time that retains its form when the oscillation is transmitted through a linear system (with constant parameters); the amplitude and phase of the oscillation are the only variables.

Second, the sine and cosine expansion of a compound signal allows one to use the symbolic method which has been developed to analyze the transmission of harmonic waves through linear networks.

For these and other reasons, the harmonic analysis is generally recognized in all branches of modern science and engineering.

When solving the second problem (approximate expansion of functions) use is made of various orthogonal sets of functions: Chebyshev, Hermite, Laguerre and Legendre polynomials, and many others. Some of these sets will be discussed in Chapter 14.

2.3. HARMONIC ANALYSIS OF PERIODIC OSCILLATIONS

When expanding a periodic oscillation $s(t)$ in a Fourier series in terms of trigonometric functions, one takes the following orthogonal set:

$$1, \cos \omega_1 t, \sin \omega_1 t, \cos 2\omega_1 t, \sin 2\omega_1 t, \ldots, \cos n\omega_1 t \sin n\omega_1 t, \ldots \qquad (2.18)$$

or

$$\ldots e^{-i2\omega_1 t}, e^{-i\omega_1 t}, 1, e^{i\omega_1 t}, e^{i2\omega_1 t}, \ldots \qquad (2.19)$$

In both cases the interval of orthogonality coincides with the period $T = 2\pi/\omega_1$ of the function $s(t)$.

The set (2.18) yields a Fourier trigonometric series, while the set (2.19), a Fourier complex series.

There is a simple connection between these two forms of the Fourier series.

First, let us use the orthogonal set (2.19). In this case the Fourier series must be written in the form

$$s(t) = \ldots c_{-2} e^{-i2\omega_1 t} + c_{-1} e^{-i\omega_1 t} + c_0 + c_1 e^{i\omega_1 t} + c_2 e^{i2\omega_1 t} + \ldots = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_1 t} \quad (2.20)$$

The Fourier coefficients c_n can easily be found by the formulas given in the preceding section.

From the formula (2.5′) it follows that

$$\| \varphi_n(t) \|^2 = \int_{-T/2}^{T/2} e^{in\omega_1 t} e^{-in\omega_1 t}\, dt = T \quad (2.21)$$

Thus, irrespective of n, the norm $\| \varphi_n \| = \sqrt{T}$. Using the formula (2.9′), we obtain

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} s(t)\, e^{-in\omega_1 t}\, dt \quad (2.22)$$

Expressions (2.21) and (2.22) take into account that $e^{-in\omega_1 t}$ is the complex conjugate of $e^{in\omega_1 t}$.

In the general case, the coefficients c_n are complex numbers. Substituting $e^{-in\omega_1 t} = \cos n\omega_1 t - i \sin n\omega_1 t$ into (2.22), we obtain

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} s(t) \cos n\omega_1 t\, dt - i \frac{1}{T} \int_{-T/2}^{T/2} s(t) \sin n\omega_1 t\, dt = c_{nc} - i c_{ns} \quad (2.23)$$

The cosine (real) and sine (imaginary) parts of the coefficients c_n are determined by the following formulas:

$$c_{nc} = \frac{1}{T} \int_{-T/2}^{T/2} s(t) \cos n\omega_1 t\, dt, \quad c_{ns} = \frac{1}{T} \int_{-T/2}^{T/2} s(t) \sin n\omega_1 t\, dt \quad (2.24)$$

It is often convenient to express the coefficients c_n in the form

$$c_n = |c_n|\, e^{i\theta_n} \quad (2.25)$$

where

$$|c_n| = \sqrt{c_{nc}^2 + c_{ns}^2} \tag{2.26}$$

$$\theta_n = -\arctan \frac{c_{ns}}{c_{nc}} \tag{2.27}$$

The modulus $|c_n|$ is an even function of n, while the argument θ_n is an odd function of n (this follows from expressions (2.24), which show that c_{nc} is an even and c_{ns} an odd function of n). The general expression (2.20) can be reduced to the form

$$s(t) = \sum_{n=-\infty}^{\infty} |c_n| e^{i(n\omega_1 t + \theta_n)} \tag{2.28}$$

Now it is an easy matter to go on to the trigonometric form of the Fourier series. Separating the two terms corresponding to any given value of $|n|$, for example, $|n| = 2$, and taking into account that $\theta_{-2} = -\theta_2$ and $|c_{-2}| = |c_2|$, we have for the sum of these terms:

$$|c_{-2}| e^{i(-2\omega_1 t + \theta_{-2})} + |c_2| e^{i(2\omega_1 t + \theta_2)}$$

$$= |c_2| [e^{-i(2\omega_1 t + \theta_2)} + e^{i(2\omega_1 t + \theta_2)}] = 2|c_2| \cos(2\omega_1 t + \theta_2) \tag{2.29}$$

From this it is seen that the series (2.28) can be written in a trigonometric form as

$$s(t) = c_0 + \sum_{n=1}^{\infty} 2|c_n| \cos(n\omega_1 t + \theta_n) \tag{2.30}$$

The meaning of doubling the Fourier coefficients c_n in a trigonometric series when $n \ll 1$ is readily seen from the vector diagram of Fig. 2.1, which corresponds to expression (2.29) in the case $n = 2$. The real-valued function $2|c_n| \times \cos(n\omega_1 t + \theta_n)$ is obtained as the sum of projections on the horizontal axis OB of two vectors of length $|c_n|$, which rotate in opposite directions at an angular velocity of $|n|\omega_1$. The vector rotating counterclockwise corresponds to positive frequency while that rotating clockwise corresponds to negative frequency. In the case of a trigonometric Fourier-series representation, the concept of "negative frequency" makes no sense. The coefficient c_0 is not doubled, since the zero-frequency component has no "double" in the spectrum of a periodic signal.

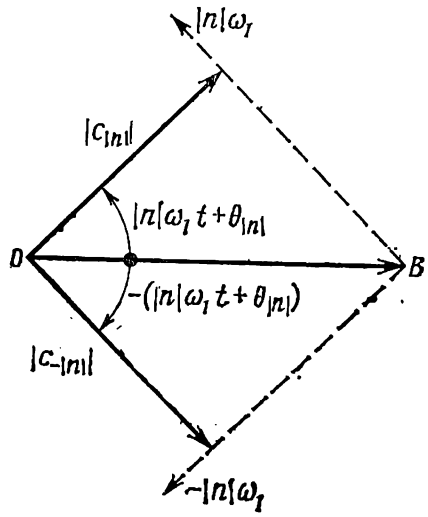


Fig. 2.1. Representation of a harmonic oscillation in the form of two complex components of positive and negative frequency

In mathematical and radio-engineering literature, expression (2.30) is often written in the following form

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega_1 t - b_n \sin n\omega_1 t)$$

$$= \frac{a_0}{2} + \sum_{n=1}^{\infty} A_n \cos (n\omega_1 t + \theta_n) \qquad (2.31)$$

Comparison between expressions (2.31) and (2.30) shows that the amplitude A_n of the nth harmonic is related to the coefficient $|c_n|$ of the series (2.28) by the relation $A_n = 2|c_n|$, while $a_n = 2c_{nc}$ and $b_n = 2c_{ns}$.

Thus, for all positive values of n (including $n = 0$)

$$\left.\begin{aligned} a_n &= \frac{2}{T}\int_{-T/2}^{T/2} s(t)\cos n\omega_1 t\, dt \\ b_n &= \frac{2}{T}\int_{-T/2}^{T/2} s(t)\sin n\omega_1 t\, dt \end{aligned}\right\} \qquad (2.32)$$

If an oscillation is an even function of t, i.e., if $s(t) = s(-t)$, only cosine terms remain in its trigonometric Fourier-series representation, since the coefficients b_n, in accordance with formula (2.32), vanish. On the other hand, in the case of an odd function $s(t)$, the coefficients a_n vanish and the series consists only of sine terms.

Two characteristics — amplitude and phase, i.e., the moduli and arguments of the complex Fourier coefficients — completely define

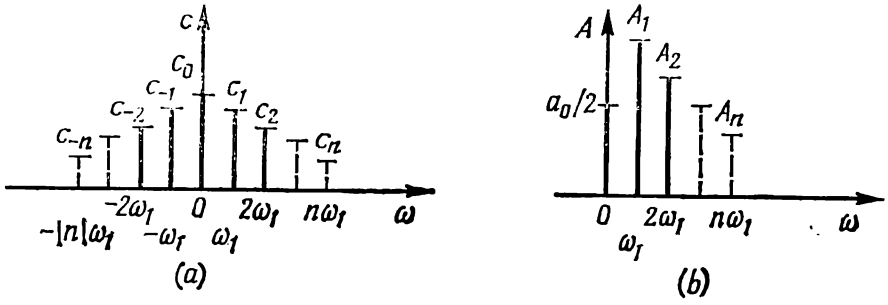


Fig. 2.2. Coefficients of (*a*) the Fourier complex and (*b*) trigonometric series for a periodic function of time

the structure of the frequency spectrum of a periodic oscillation. The "width" of the spectrum can be visualized by representing graphically the spectrum of amplitudes. As an example, Fig. 2.2*a*

shows the spectrum of coefficients $|c_n|$, and Fig. 2.2*b*, that of amplitudes $A_n = 2|c_n|$ of one and the same periodic oscillation. To characterize the spectrum exhaustively, such constructions must be supplemented with the data on the phase angles of the individual harmonics.

The spectrum of a periodic function is called *line* or *discrete*, because it consists of separate lines corresponding to discrete frequencies: 0, ω_1, $\omega_2 = 2\omega_1$, $\omega_3 = 3\omega_1$, etc.

The use of the Fourier series for the harmonic analysis of complex periodic signals, in conjunction with the superposition principle, is an effective means for the study of the effect of linear systems on the propagation of signals. It should be noted, however, that the determination of an oscillation at the output of a system by the sum of harmonics with specified amplitudes and phases is not an easy task, particularly when the Fourier series representing the oscillation does not converge rapidly. The signals radio engineering makes the most wide use of do not satisfy this condition, and for an adequate reproduction of their form, a great number of harmonics have generally to be added together. For this reason, it is to be thought that in the case of complex periodic signals, the Fourier-series representation is more convenient to use for the analysis of signals rather than for their synthesis.

2.4. SPECTRA OF SIMPLE PERIODIC OSCILLATIONS

Let us consider some periodic oscillations that are widely used in various radio engineering systems.

2.4-1. Square-Wave Oscillation (Fig. 2.3)

Such an oscillation, often referred to as *meander* (from the Greek *maiandros*, the ancient name of the Menderes River noted for its

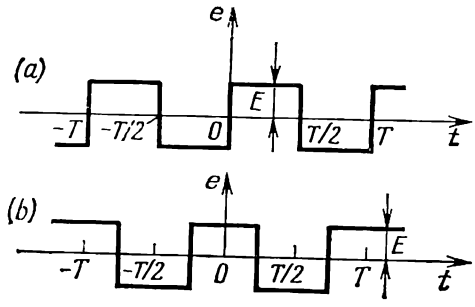


Fig. 2.3. Square-wave periodic oscillation (meander)

winding course), is widely used in measuring instruments and systems.

If the origin of time is selected as shown in Fig. 2.3*a*, the function describing such an oscillation will be odd, but if it is chosen as in

Fig. 2.3*b*, the function will be even. Using formulas (2.24), we find for the oscillation shown in Fig. 2.3*a*

$$c_{nc}=0$$

$$c_{ns}=\frac{1}{T}\int_{-T/2}^{T/2} e(t)\sin n\omega_1 t\,dt$$

$$=\frac{E}{T}\left[\int_{-T/2}^{0}(-1)\sin n\omega_1 t\,dt+\int_{0}^{T/2}\sin n\omega_1 t\,dt\right]=\frac{2E}{Tn\omega_1}\left(1-\cos\frac{n\omega_1 T}{2}\right)$$

Taking into account that $T\omega_1 = 2\pi$, we have

$$c_{ns}=\frac{E}{\pi n}(1-\cos n\pi)=\begin{cases}0 & (n=0,2,4,\ldots)\\ 2E/n\pi & (n=1,3,5,\ldots)\end{cases}$$

The epoch angles θ_n, in accordance with (2.27), are equal to $-\pi/2$ for all the harmonics.

Let us write a Fourier trigonometric series

$$e(t)=\sum_{n=1,3,5,\ldots}^{\infty} 2|c_{ns}|\cos\left(n\omega_1 t-\frac{\pi}{2}\right)$$

$$=\frac{4E}{\pi}\left(\sin\omega_1 t+\frac{1}{3}\sin 3\omega_1 t+\frac{1}{5}\sin 5\omega_1 t+\ldots\right) \qquad (2.33)$$

The spectrum of the coefficients $|c_n|$ of a Fourier complex series is shown in Fig. 2.4*a* while that of the trigonometric series is presented in Fig. 2.4*b*.

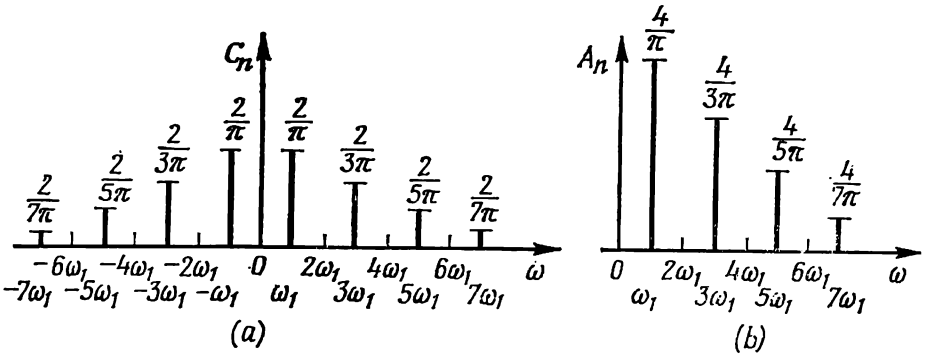

C_n
$\frac{2}{7\pi}$ $\frac{2}{5\pi}$ $\frac{2}{3\pi}$ $\frac{2}{\pi}$ $\frac{2}{\pi}$ $\frac{2}{3\pi}$ $\frac{2}{5\pi}$ $\frac{2}{7\pi}$
$-6\omega_1$ $-4\omega_1$ $-2\omega_1$ 0 $2\omega_1$ $4\omega_1$ $6\omega_1$ ω
$-7\omega_1$ $-5\omega_1$ $-3\omega_1$ $-\omega_1$ ω_1 $3\omega_1$ $5\omega_1$ $7\omega_1$
(a)
A_n
$\frac{4}{\pi}$ $\frac{4}{3\pi}$ $\frac{4}{5\pi}$ $\frac{4}{7\pi}$
0 $2\omega_1$ $4\omega_1$ $6\omega_1$ ω
ω_1 $3\omega_1$ $5\omega_1$ $7\omega_1$
(b)

Fig. 2.4. Coefficients of (*a*) the Fourier complex and (*b*) trigonometric series for the oscillation shown in Fig. 2.3

If the origin of time is taken to lie at the centre of a pulse (Fig. 2.3*b*) the function will be even with respect to t, in which case

$$e(t)=\frac{4E}{\pi}\left(\cos\omega_1 t-\frac{1}{3}\cos 3\omega_1 t+\frac{1}{5}\cos 5\omega_1 t-\ldots\right) \qquad (2.34)$$

The diagrams of the first ($n = 1$) and the third ($n = 3$) harmonic and of their sum are shown in Fig. 2.5*a*. Figure 2.5*b* shows this sum supplemented with the fifth ($n = 5$) harmonic, and in the diagram of Fig. 2.5*c*, the seventh ($n = 7$) harmonic is added.

As the number of the harmonics being added together increases, the sum of the series approaches the function $e(t)$ everywhere except for

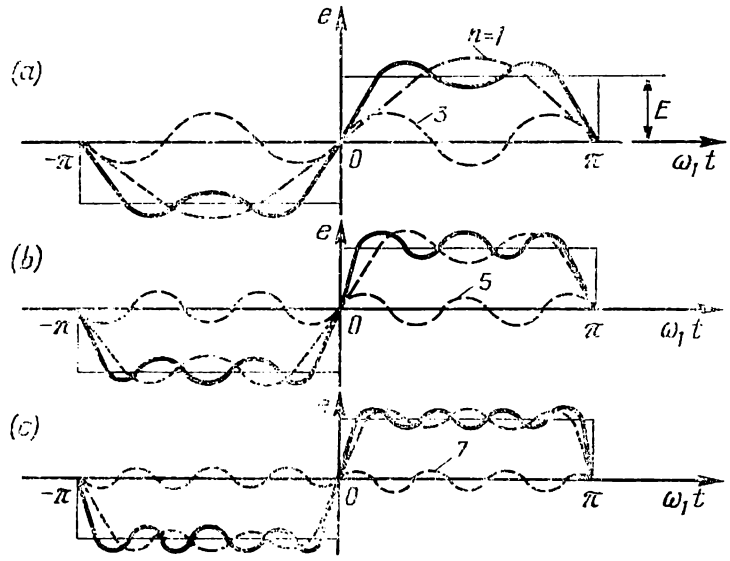


Fig. 2.5. Addition of (*a*) the 1st and 3rd harmonics, (*b*) the 1st, 3rd and 5th harmonics, and (*c*) the 1st, 3rd, 5th, and 7th harmonics of the oscillation shown in Fig. 2.3

the discontinuities of the function, where an "overshoot" is formed. In the case $n \to \infty$, the magnitude of overshooting is equal to $1.18E$, i.e., the sum of the series is 18% in excess of the true value of the given function. This convergence defect (nonuniform convergence near the discontinuities) is known in mathematics as *Gibbs phenomenon*. In spite of the fact that in the given case the Fourier series does not converge to the function $e(t)$ at the discontinuities, the series converges in mean, since in the case $n \to \infty$ the overshoots are infinitely narrow and do not add to the magnitude of the integral (2.13).

2.4-2. Saw-Tooth Oscillation (Fig. 2.6)

Such functions are often dealt with in the sweep circuits of oscilloscopes. Since this function is odd, the respective Fourier series comprises sinusoidal terms only. The Fourier coefficients can easily

be found by using formulas (2.24) through (2.31). Let these computations be omitted and the final expression for the series written as

$$e(t) = \frac{2E}{\pi}\left(\sin\omega_1 t - \frac{1}{2}\sin 2\omega_1 t + \frac{1}{3}\sin 3\omega_1 t - \frac{1}{4}\sin 4\omega_1 t + \ldots\right) \quad (2.35)$$

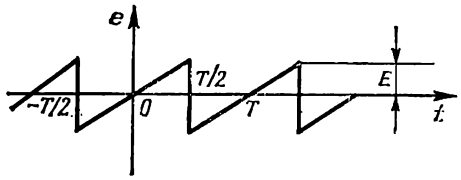


Fig. 2.6. Saw-tooth periodic oscillation

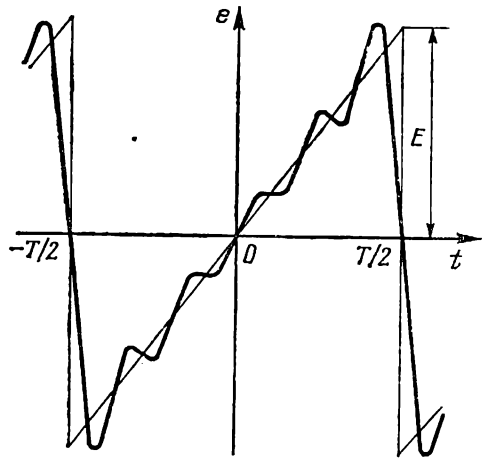


Fig. 2.7. Sum of the first five harmonics of the oscillation shown in Fig. 2.6

It is clear that the amplitudes of the harmonics diminish as $1/n$, where $n = 1, 2, 3, \ldots$ Figure 2.7 shows (to an enlarged scale) the sum of the first five harmonics.

2.4-3. Train of Unipolar Triangular Pulses (Fig. 2.8)

For this function, the Fourier series has the following form:

$$e(t) = \frac{E}{\pi}\left[\frac{\pi}{2} - \frac{4}{\pi}\left(\cos\omega_1 t + \frac{1}{3^2}\cos 3\omega_1 t + \frac{1}{5^2}\cos 5\omega_1 t + \ldots\right)\right] \quad (2.36)$$

Figure 2.8 illustrates the sum of the first three members of this series. Note that in this case the amplitudes of the harmonics dimi-

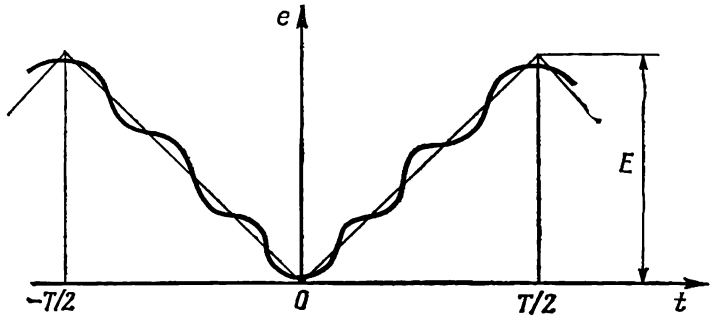


Fig. 2.8. Sum of the first three harmonics of a periodic function

nish at a higher rate than in the preceding examples. This is explained by the absence of discontinuities in the function.

2.4-4. Train of Unipolar Rectangular Pulses (Fig. 2.9)

Using expressions (2.32), we can find the mean value (direct-current component)

$$\frac{a_0}{2} = \frac{1}{T} \int_{-\tau_p/2}^{\tau_p/2} e(t)\, dt = \frac{\tau_p}{T} E \tag{2.37}$$

and the coefficient of the nth harmonic

$$a_n = \frac{2}{T} \int_{-\tau_p/2}^{\tau_p/2} e(t) \cos n\omega_1 t\, dt = \frac{2E}{\pi n} \sin \frac{n\omega_1 \tau_p}{2} \tag{2.38}$$

Since the function $e(t)$ is even, $b_n = 0$ and $A_n = a_n$. Thus,

$$e(t) = E\left(\frac{\tau_p}{T} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin \frac{n\omega_1 \tau_p}{2} \cos n\omega_1 t\right) \tag{2.39}$$

The quantity $N = T/\tau_p$ is called *pulse period-to-pulse width ratio* (or reciprocal *duty factor*). When N is large, the signal spectrum includes a great number of harmonics with slowly diminishing amplitudes (Fig. 2.10). The distance between the spectrum lines is very small, and the amplitudes of the adjacent harmonics are

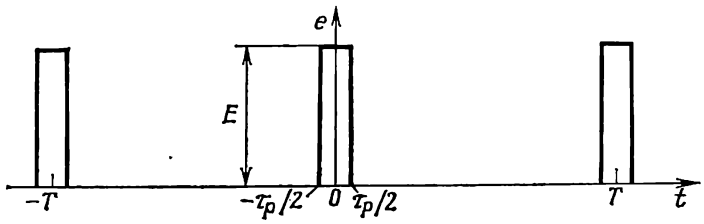


Fig. 2.9. Periodic train of rectangular pulses with a large reciprocal duty factor

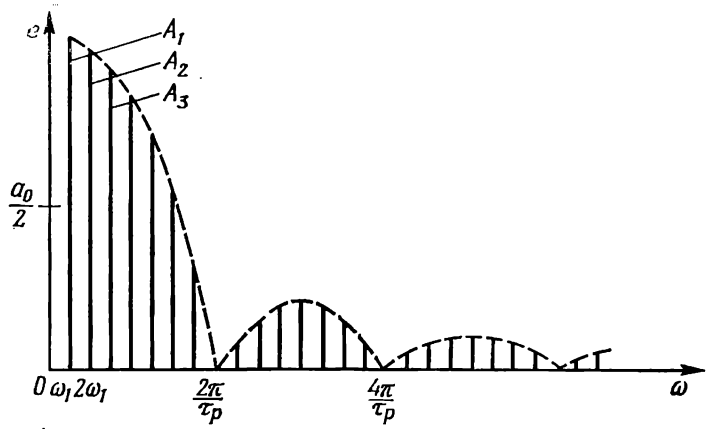


Fig. 2.10. Spectrum of the pulse train shown in Fig. 2.9

close in magnitude. This is clear from formula (2.38), which in this case is preferably written as

$$|a_n| = A_n = \frac{2E}{\pi n}\left|\sin\left(n\pi\frac{\tau_p}{T}\right)\right|$$

For small n we may assume that

$$A_n \approx \frac{2E}{\pi n}\frac{n\pi\tau_p}{T} = E\frac{2\tau_p}{T} \qquad (2.40)$$

The direct-current component, $a_0/2 = E\tau_p/T$, is half the amplitude of the first harmonic. If one had to construct a diagram of the spectrum of the coefficients $|c_n|$, one would have found that the value of c_0 is approximately equal to $|c_1|$.

2.5. POWER DISTRIBUTION IN THE SPECTRUM OF A PERIODIC OSCILLATION

Let an oscillation $s(t)$ (current, voltage) be a compound periodic function of time with a period T.

The total energy of such an oscillation over the interval from $t = -\infty$ to $t = \infty$ is infinitely large. Of special interest is the

average power of the oscillation and the power distribution among its individual harmonics. It is clear that the average power over the entire time axis coincides with the average power over a single period T. Therefore, we may use formula (2.17), taking the coefficients c_n to be those of the series (2.20), the interval of orthogonality, $t_2 - t_1$, to be the period T, and the norm $\| \varphi_n \|$, the quantity $\sqrt{T}$ [see formula (2.21)].

Thus, the average power of a periodic oscillation is

$$\overline{s^2(t)} = \frac{1}{T} \sum_{n=-\infty}^{\infty} |c_n|^2 T = \sum_{n=-\infty}^{\infty} |c_n|^2 \qquad (2.41)$$

Using the trigonometric form of the Fourier series and taking into account that $c_0 = a_0/2$ and $|c_n| = A_n/2$, we get

$$\overline{s^2(t)} = \left(\frac{a_0}{2}\right)^2 + 2 \sum_{n=1}^{\infty} \left(\frac{A_n}{2}\right)^2 = \left(\frac{a_0}{2}\right)^2 + \frac{1}{2} \sum_{n=1}^{\infty} A_n^2 \qquad (2.42)$$

If $s(t)$ is the current $i(t)$ flowing through a resistor r, the average power dissipated in the resistor is

$$P = \overline{ri^2(t)} = r\left(I_0^2 + \frac{I_1^2}{2} + \frac{I_2^2}{2} + \ldots\right)$$

where $I_0 = a_0/2$ is the *d-c* component and $I_n = A_n$, the amplitude of the nth harmonic of the current $i(t)$.

Thus, the total average power of a periodic oscillation is equal to the sum of the average powers dissipated separately by the *d-c* component I_0 and the harmonics (with amplitudes I_1, I_2, etc.). This means that the *average power is independent of the phases of the individual harmonics.*

2.6. HARMONIC ANALYSIS OF NONPERIODIC OSCILLATIONS

The harmonic analysis of periodic oscillations discussed in Section 2.3 can easily be applied to nonperiodic oscillations. Let such

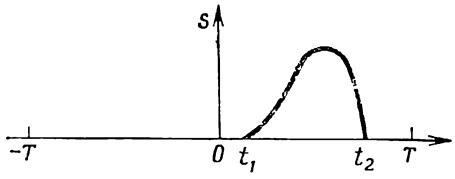


Fig. 2.11. Solitary pulse

an oscillation $s(t)$ be given as some function which is nonzero in an interval (t_1, t_2) (Fig. 2.11).

By selecting an arbitrary time interval T including the interval (t_1, t_2), we may represent the given oscillation as a Fourier series

$$s(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_1 t} \quad (0 < t < T) \tag{2.43}$$

where $\omega_1 = 2\pi/T$ and the coefficients c_n, according to formula (2.22), are

$$c_n = \frac{1}{T} \int_{t_1}^{t_2} s(t)\, e^{-in\omega_1 t}\, dt \tag{2.44}$$

Substituting (2.44) into (2.43) and taking into consideration that $T = 2\pi/\omega_1$, we obtain

$$s(t) = \sum_{n=-\infty}^{\infty} \left(\int_{t_1}^{t_2} s(x)\, e^{-in\omega_1 x}\, dx \right) e^{in\omega_1 t} \frac{\omega_1}{2\pi} \quad (0 < t < T) \tag{2.45}$$

Outside the interval $(0, T)$ the series (2.43) defines the function $s(t) = s(t \pm kT)$, where k is an integer, i.e., it defines a periodic function obtained by repeating $s(t)$ to the right and left with a period T. For the function to vanish outside the interval $(0, T)$ the value of T must be infinitely large. But the larger the period T, the smaller the coefficients c_n. When T approaches infinity, we obtain in the limit infinitely low amplitudes of the harmonics whose sum represents the original nonperiodic function $s(t)$ defined in the interval $t_1 < t < t_2$ (Fig. 2.11). The number of harmonics entering into the Fourier series will in this case be infinitely large, because as $T \to \infty$, the fundamental frequency $\omega_1 = 2\pi/T \to 0$. In other words, the distance between the spectral lines, which is equal to the fundamental frequency ω_1, becomes infinitely small, and the spectrum becomes continuous (see Fig. 2.2).

Therefore, in expression (2.45) one can substitute $d\omega$ for ω_1 and the running frequency ω for $n\omega_1$, and perform integration instead of summation.

Thus, we come to the double Fourier integral

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \left[\int_{t_1}^{t_2} s(x)\, e^{-i\omega x}\, dx \right] d\omega \tag{2.46}$$

The internal integral (which is a function of ω)

$$S(\omega) = \int_{t_1}^{t_2} s(t)\, e^{-i\omega t}\, dt \tag{2.47}$$

is called *the spectral density* or *spectral characteristic of the function* $s(t)$.

In the general case where the limits t_1 and t_2 are not specified, spectral density is written in the form

$$\mathbf{S}(\omega) = \int_{-\infty}^{\infty} s(t)\, e^{-i\omega t}\, dt \tag{2.48}$$

Substituting (2.48) into expression (2.46), we obtain

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, e^{i\omega t}\, d\omega \tag{2.49}$$

Expressions (2.48) and (2.49) are known respectively as *Fourier transform* and *inverse Fourier transform*.

The only difference between expressions (2.48) and (2.22) is in the factor $1/T$. Consequently, the spectral density $\mathbf{S}(\omega)$ features all the basic properties of the coefficients c_n of a Fourier complex series. By analogy with (2.23) and (2.24), we may write

$$\mathbf{S}(\omega) = A(\omega) - iB(\omega) = S(\omega)\, e^{i\theta(\omega)} \tag{2.50}$$

where

$$\left.\begin{aligned} A(\omega) &= \int_{-\infty}^{\infty} s(t) \cos \omega t\, dt \\ B(\omega) &= \int_{-\infty}^{\infty} s(t) \sin \omega t\, dt \end{aligned}\right\} \tag{2.51}$$

The modulus and argument of the spectral density function are defined by the expressions

$$S(\omega) = \sqrt{[A(\omega)]^2 + [B(\omega)]^2} \tag{2.52}$$

$$\theta(\omega) = -\arctan [B(\omega)/A(\omega)] \tag{2.53}$$

The first of these expressions may be regarded as the *amplitude-frequency characteristic* and the second, as the *phase-frequency characteristic* of the continuous spectrum of a nonperiodic oscillation $s(t)$. As in the case of the Fourier series, $S(\omega)$ is an even and $\theta(\omega)$, an odd function of frequency ω.

On the basis of formula (2.50), expression (2.49) can easily be reduced to a trigonometric form. We have

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)\, e^{i(\omega t + \theta)}\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \cos(\omega t + \theta)\, d\omega + i \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \sin(\omega t + \theta)\, d\omega$$

Since the modulus is an even function and the phase, an odd one, it follows that the integrand in the first integral is even and in the second integral, odd with respect to ω. Consequently, the second integral is zero and finally,

$$s(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\cos(\omega t+\theta)\,d\omega=\frac{1}{\pi}\int_{0}^{\infty} S(\omega)\cos(\omega t+\theta)\,d\omega \quad (2.54)$$

It is advisable to carry out the transformation from the complex form (2.49) to the trigonometric form (2.54) at the end of an analysis, all the intermediate computations with the use of the Fourier integral being easier and more convenient to perform on the basis of the complex equation (2.49).

From a comparison of expressions (2.49) and (2.20) we may see that the quantity

$$\frac{1}{2\pi}\mathbf{S}(\omega)\,d\omega=\mathbf{S}(\omega)\,df$$

has the meaning of the coefficient c_n (infinitesimal) of a Fourier complex series at a frequency $\omega = 2\pi f$.

Correspondingly, from a comparison of expressions (2.54) and (2.31) it is clear that the quantity $\frac{1}{\pi}S(\omega)\,d\omega = 2S(\omega)df$ has the meaning of the amplitude A_n (infinitesimal) of a harmonic of the frequency $\omega = 2\pi f$.

These comparisons explain the meaning of the term "spectral density". The quantity $2S(\omega)$ is the oscillation amplitude per Hz in the infinitely narrow frequency band including the given frequency ω.

On the basis of the above discussion, we can easily determine the relationship between the spectra of a single pulse and a periodic train of pulses.

Let $\mathbf{S}_1(\omega)$ be the spectrum of a single pulse $s_1(t)$ and T, the repetition period. As said above, the spectral density $\mathbf{S}_1(\omega)$ [see (2.48)] differs from the coefficient c_n of the Fourier series for a periodic sequence merely in the factor $1/T$ [see (2.22)]. From this it follows that when the pulse $s_1(t)$ is repeated with a period T, the coefficients c_n of the Fourier series for the periodic sequence thus obtained are given by

$$c_n = \mathbf{S}_1(\omega)/T \quad (2.55)$$

where the argument ω of the spectral density $\mathbf{S}_1(\omega)$ should be taken equal to the frequency $n\omega_1$ of the corresponding harmonic.

Thus,

$$\mathbf{A}_n=2c_n=\frac{2\mathbf{S}_1(n\omega_1)}{T_1}=\frac{\omega_1}{\pi}\mathbf{S}_1(n\omega_1) \quad (2.56)$$

Consequently, *the modulus of the spectral density of a single pulse and the envelope of the line spectrum of the periodic train of pulses obtained by repetition of the given pulse coincide in form and differ only in scale.*

2.7. SOME PROPERTIES OF FOURIER TRANSFORMS

A unique relation exists between an oscillation $s(t)$ and its spectrum $S(\omega)$. For practical purposes, it is important to find the connection between a transformation of an oscillation and the respective change in its spectrum. From among the numerous possible transformations of an oscillation let us consider the most important and frequently used ones—time shift of an oscillation, change of time scale, frequency shift of an oscillation spectrum, differentiation and integration of an oscillation. In addition, we shall consider the sum, product, and convolution of two oscillations and also the interchangeability of ω and t in Fourier transforms.

2.7-1. Time Shift of an Oscillation

Assume that an arbitrary oscillation $s_1(t)$ having a spectral density $S_1(\omega)$ exists in a time interval from t_1 to t_2. By delaying this oscillation (without distorting its form) for a time t_0, we obtain a new function of time

$$s_2(t) = s_1(t - t_0)$$

existing within an interval from $t_1 + t_0$ to $t_2 + t_0$.

The spectral density of the signal $s_2(t)$, in accordance with (2.48), is given by

$$S_2(\omega) = \int_{-\infty}^{\infty} s_2(t)\, e^{-i\omega t}\, dt = \int_{t_1+t_0}^{t_2+t_0} s_1(t-t_0)\, e^{-i\omega t}\, dt$$

Introducing a new integration variable $\tau = t - t_0$, we have

$$S_2(\omega) = e^{-i\omega t_0} \int_{-\infty}^{\infty} s_1(\tau)\, e^{-i\omega\tau}\, d\tau = e^{-i\omega t_0} S_1(\omega) \qquad (2.57)$$

From this relation it is clear that a $\pm t_0$ time shift of the function $s(t)$ leads to a $\pm\omega t_0$ change in the phase characteristic of the spectrum $S(\omega)$. The reverse statement is evident: if all the components of the spectrum of the function $s(t)$ are subjected to a phase shift $\theta(\omega) = \pm\omega t_0$, linearly related to the frequency ω, the function will be shifted in time by $\pm t_0$.

The amplitude-frequency characteristic of the spectrum (i.e., the modulus of the spectral density function) is independent of the position of the oscillation on the time axis.

2.7-2. Change of Time Scale

Let an oscillation $s_1(t)$ (shown in Fig. 2.12 by a solid line) be subjected to compression in time. The new, compressed oscillation $s_2(t)$ (shown by a broken line in Fig. 2.12) is related to the initial oscillation $s_1(t)$ by the relation

$$s_2(t) = s_1(nt) \quad (n > 1)$$

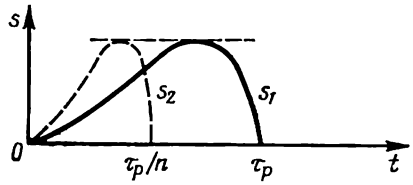


Fig. 2.12. Compression of a signal with its shape and amplitude being preserved

The duration of the pulse $s_2(t)$ is a factor of n shorter than that of the initial pulse and is equal to τ_p/n. The spectral density of the compressed pulse is

$$S_2(\omega) = \int_0^{\tau_p/n} s_2(t)\, e^{-i\omega t}\, dt = \int_0^{\tau_p/n} s_1(nt)\, e^{-i\omega t}\, dt$$

Introducing a new integration variable $\tau = nt$, we obtain

$$S_2(\omega) = \frac{1}{n} \int_0^{\tau_p} s_1(\tau)\, e^{-i(\omega/n)\tau}\, d\tau$$

But the integral on the right-hand side of this equation is none other than the spectral density of the initial oscillation $s_1(t)$ at a frequency ω/n, i.e., it is equal to $S_1(\omega/n)$.

So,

$$S_2(\omega) = (1/n)\, S_1(\omega/n)$$

Thus, when compressing the oscillation by a factor of n on the time axis, its frequency spectrum widens in proportion on the frequency axis. In this case the modulus of the spectral density function is reduced by a factor of n. It is evident that when stretching the oscillation in time (i.e., when $n < 1$), the frequency spectrum narrows and the modulus of the spectral density function increases.

2.7-3. Shift of an Oscillation Spectrum

Use the Fourier transform (2.48) for a product $s(t)\cos(\omega_0 t + \theta_0)$:

$$\int_{-\infty}^{\infty} s(t)\cos(\omega_0 t+\theta_0)\, e^{-i\omega t}\, dt$$

$$= \int_{-\infty}^{\infty} s(t)\left[\frac{1}{2} e^{i(\omega_0 t+\theta_0)} + \frac{1}{2} e^{-i(\omega_0 t+\theta_0)}\right] e^{-i\omega t}\, dt$$

$$= \frac{e^{-i\theta_0}}{2} \int_{-\infty}^{\infty} s(t)\, e^{-i(\omega-\omega_0)t}\, dt + \frac{e^{-i\theta_0}}{2} \int_{-\infty}^{\infty} s(t)\, e^{-i(\omega+\omega_0)t}\, dt$$

The first integral on the right-hand side is none other than the spectral density of the function $s(t)$ at a frequency $\omega - \omega_0$, while the second integral, the same at a frequency $\omega + \omega_0$. Therefore, the above expression can be written in the form

$$\int_{-\infty}^{\infty} s(t)\cos(\omega_0 t+\theta_0)\, e^{-i\omega t}\, dt = \frac{1}{2}\left[e^{i\theta_0}\mathbf{S}(\omega-\omega_0) + e^{-i\theta_0}\mathbf{S}(\omega+\omega_0)\right] \qquad (2.58)$$

where $\mathbf{S}(\omega)$ is the spectral density of the oscillation $s(t)$.

From expression (2.58) it follows that splitting the spectrum $\mathbf{S}(\omega)$ into two parts shifted respectively by $+\omega_0$ and $-\omega_0$ is equivalent to multiplying the function $s(t)$ by a harmonic oscillation $\cos\omega_0 t$ (with $\theta_0 = 0$).

This proposition is considered in greater detail in Chapter 3.

2.7-4. Differentiation and Integration of an Oscillation

Let us omit rigorous proofs and restrict ourselves to simple reasoning. The differentiation of an oscillation $s_1(t)$ may be considered a term-by-term differentiation of all the harmonics entering into its spectrum. In the general case, an ω-frequency component of an oscillation $s_1(t)$ may be expressed as

$$\left[\frac{1}{2\pi}\mathbf{S}_1(\omega)\, d\omega\right] e^{i\omega t}$$

The quantity in the square brackets may be regarded as the amplitude of the oscillation in the frequency band $d\omega$ [cf (2.49)].

Differentiation with respect to time t yields

$$i\omega\left[\frac{1}{2\pi}\mathbf{S}_1(\omega)\, d\omega\right] e^{i\omega t}$$

Hence, the spectral density of the derivative $ds_1(t)/dt$ is

$$\mathbf{S}_2(\omega) = i\omega \mathbf{S}_1(\omega) \tag{2.59}$$

Similarly, the spectral density of the integral $\int s_1(t)\,dt$ is

$$\mathbf{S}_2(\omega) = (1/i\omega)\,\mathbf{S}_1(\omega) \tag{2.60}$$

2.7-5. Addition of Oscillations

Since the Fourier transform determining the spectral density of a given function of time is a linear operation, it is clear that when adding oscillations $s_1(t)$, $s_2(t)$ . . ., having spectra $\mathbf{S}_1(\omega)$, $\mathbf{S}_2(\omega)$, ..., the spectrum $\mathbf{S}(\omega)$ of the resulting oscillation $s(t) = s_1(t) + s_2(t) + \ldots$ is equal to the sum $\mathbf{S}_1(\omega) + S_2(\omega) + \ldots$

2.7-6. Product of Two Oscillations

Let an oscillation $s(t)$ be a product of two functions of time, $f(t)$ and $g(t)$.

Using the general formula (2.48), we find the spectrum of the oscillation $s(t)$:

$$\mathbf{S}(\omega) = \int_{-\infty}^{\infty} s(t)\, e^{-i\omega t}\, dt = \int_{-\infty}^{\infty} f(t)\, g(t)\, e^{-i\omega t}\, dt \tag{2.61}$$

Each function $f(t)$ and $g(t)$, can be represented in the form of a Fourier integral:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{F}(\omega)\, e^{i\omega t}\, d\omega; \quad g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{G}(\omega)\, e^{i\omega t}\, d\omega$$

Substituting the second of these integrals into equation (2.61), we obtain

$$\mathbf{S}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\, dt \int_{-\infty}^{\infty} \mathbf{G}(x)\, e^{ixt}\, dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{G}(x) \left[\int_{-\infty}^{\infty} f(t)\, e^{-i(\omega - x)t}\, dt \right] dx$$

The integral in the square brackets represents the spectral density of the function $f(t)$ at a frequency $\omega - x$, i.e., $\mathbf{F}(\omega - x)$. Consequently

$$\mathbf{S}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{G}(x)\, \mathbf{F}(\omega - x)\, dx \tag{2.62}$$

Thus, the spectrum of the product of two functions of time, $f(t)$ and $g(t)$, is equal (with a coefficient $1/2\pi$) to the convolution of their spectra $\mathbf{F}(\omega)$ and $\mathbf{G}(\omega)$.

In the particular case $\omega = 0$, expressions (2.61) and (2.62) yield the following equality

$$\int_{-\infty}^{\infty} f(t)\, g(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{G}(x)\, \mathbf{F}(-x)\, dx$$

Substituting ω for x in the last expression, we obtain

$$\int_{-\infty}^{\infty} f(t)\, g(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{G}(\omega)\, \mathbf{F}(-\omega)\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{G}(\omega)\, \mathbf{F}^*(\omega)\, d\omega \qquad (2.63)$$

where $\mathbf{F}^*(\omega) = \mathbf{F}(-\omega)$ is the complex conjugate of the function $\mathbf{F}(\omega)$.

In quite a similar manner we may prove that the product of two spectra, $\mathbf{F}(\omega)\, \mathbf{G}(\omega) = \mathbf{S}(\omega)$, corresponds to a function of time, $s(t)$, which is the convolution of two functions $f(t)$ and $g(t)$:

$$s(t) = \int_{-\infty}^{\infty} f(y)\, g(t-y)\, dy = \int_{-\infty}^{\infty} f(t-y)\, g(y)\, dy$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{F}(\omega)\, \mathbf{G}(\omega)\, e^{i\omega t}\, d\omega \qquad (2.64)$$

The last expression is widely used in the analysis of the transmission of signals through linear circuits. In this case the functions of time, $f(t)$ and $g(t)$, have the meaning of the input signal and the impulse response of a circuit, respectively (see Section 6.3), while $\mathbf{F}(\omega)$ and $\mathbf{G}(\omega)$, that of the spectral density of the signal and the transfer function of the circuit.

2.7-7. Interchangeability of ω and t in Fourier Transforms

Let us consider the general expression (2.48) and clarify the character of the function $\mathbf{S}(\omega)$ for different functions $s(t)$.

(a) Let $s(t)$ be an even function with respect to t. After rewriting expression (2.48) as

$$\mathbf{S}(\omega) = \int_{-\infty}^{\infty} s(t) \cos \omega t\, dt - i \int_{-\infty}^{\infty} s(t) \sin \omega t\, dt$$

we can see that if $s(t)$ is an even function, the second integral vanishes, since the product $s(t) \sin \omega t$ is an odd function with respect to t, while the limits of integration are symmetrical.

Thus, when $s(t)$ is even with respect to t, the function $S(\omega)$ is determined by the first integral and is an even real-valued function with respect to ω.

(b) If $s(t)$ is odd with respect to t, the first integral vanishes, and

$$S(\omega) = -i \int_{-\infty}^{\infty} s(t) \sin(\omega) t \, dt$$

In this case $S(\omega)$ is an odd and purely imaginary function of ω.

(c) Finally, if $s(t)$ is neither an even nor an odd function with respect to t, it can be expanded into two functions: an even function $s_1(t)$ and an odd function $s_2(t)$.

In this case $S(\omega)$ is a complex function, its real part being even and its imaginary part, odd with respect to ω.

It follows from (a) that in the case of an even function $s(t)$ the sign of t in the inverse Fourier transform [Formula (2.49)] can be arbitrarily chosen. Let us take the minus sign and write formula (2.49) in the form

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{-i\omega t} d\omega$$

Now we substitute in the last integral the integration variable t for ω and the parameter ω for t. In this case the left-hand side should be written as a function of the argument ω:

$$s(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(t) e^{-i\omega t} dt$$

But the integral in the last expression may be regarded as the spectral density of a new function $S(t)$ obtained by replacing ω by t in the spectral density function of the signal $s(t)$.

Let us denote this spectral density by $S'(\omega)$. Then

$$S'(\omega) = 2\pi s(\omega) \tag{2.65}$$

This shows that the variables ω and t in Fourier transforms are *interchangeable*: if the spectrum $S(\omega)$ corresponds to an even oscillation $s(t)$, the spectrum $2\pi s(\omega)$ corresponds to an oscillation $S(t)$.

The application of this rule is illustrated in Sec. 2.9-3.

2.8. ENERGY DISTRIBUTION IN THE SPECTRUM OF A NONPERIODIC OSCILLATION

In order to obtain an expression similar to (2.42), we may proceed in two ways: to pass to the limit $T \to \infty$ in (2.42) or to use the results of the preceding section.

Let us consider the second way. For this purpose, we shall use expression (2.63).

If $f(t)$ and $g(t)$ represent one and the same oscillation

$$f(t) = g(t) = s(t)$$

the integral

$$\int_{-\infty}^{\infty} f(t)\, g(t)\, dt = \int_{-\infty}^{\infty} s^2(t)\, dt = E$$

represents the total energy of the oscillation $s(t)$, while the product of the spectral densities

$$\mathbf{G}(\omega)\, \mathbf{F}^*(\omega) = \mathbf{S}(\omega)\, \mathbf{S}^*(\omega) = |\mathbf{S}(\omega)|^2 = S^2(\omega)$$

where $\mathbf{S}(\omega)$ is the spectrum of the oscillation $s(t)$ and $S(\omega)$ is the modulus of this spectrum.

Thus, in accordance with (2.63), we come to the final result

$$E = \int_{-\infty}^{\infty} s^2(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathbf{S}(\omega)|^2\, d\omega = \frac{1}{\pi} \int_{0}^{\infty} S^2(\omega)\, d\omega \qquad (2.66)$$

This important relation between the energy of an oscillation (across a resistance of 1 ohm) and the modulus of its spectral density is known as *Parseval's equation.*

There is an important difference between expressions (2.42) and (2.66). In Section 2.5 we have dealt with the average power of a periodic oscillation. The averaging have been effected by dividing the energy of a "segment" of the oscillation per period by T. On the other hand, in the case of a nonperiodic oscillation of a finite duration, the averaging of energy over an infinite period gives zero; therefore, the average power of such an oscillation is equal to zero.

From expression (2.66) it is evident that the quantity $|\mathbf{S}(\omega)|^2$ which has the meaning of energy per unit frequency interval may be regarded as the *spectral density of the energy of the oscillation.*

2.9. EXAMPLES OF DETERMINING SPECTRA OF NONPERIODIC OSCILLATIONS

The main task of this Section is to illustrate the properties of Fourier transforms, discussed in the preceding sections, by means of practical examples.

2.9-1. Square Pulse

The simple signal defined by the expression

$$s_1(t) = \begin{cases} A \text{ if } -\tau_p/2 \leqslant t \leqslant \tau_p/2 \\ 0 \text{ if } t < -\tau_p/2 \text{ or } t > \tau_p/2 \end{cases} \tag{2.67}$$

and illustrated in Fig. 2.13 is widely used both in radio engineering and in the theory of signals and circuits. By employing formula (2.48), we find the spectral density (Fig. 2.14)

$$\mathbf{S}_1(\omega) = A \int_{-\tau_p/2}^{\tau_p/2} e^{-i\omega t}\, dt = \frac{A}{-i\omega}\left(e^{-i\omega\tau_p/2} - e^{i\omega\tau_p/2}\right)$$

$$= \frac{2A}{\omega} \sin\frac{\omega\tau_p}{2} = A\tau_p \left[\frac{\sin(\omega\tau_p/2)}{\omega\tau_p/2}\right] \tag{2.68}$$

Note that the product $A\tau_p$ is equal to the pulse area and determines the value of the pulse spectral density at $\omega = 0$, i.e. $\mathbf{S}_1(0) = A\tau_p$.

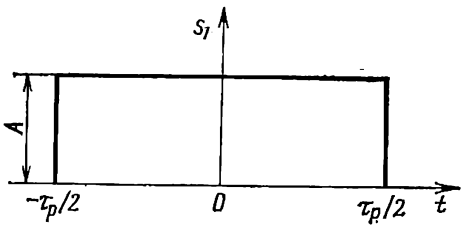


Fig. 2.13. Square pulse

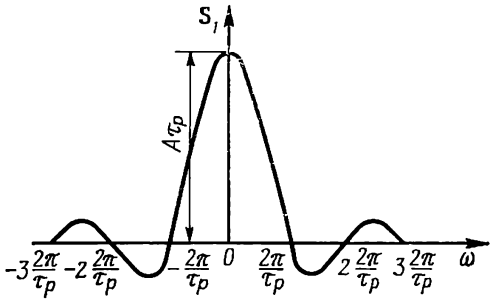


Fig. 2.14. Spectral density of a square pulse

This conclusion may be applied to a pulse of an arbitrary shape. Indeed, it follows from general equation (2.48) that

$$\mathbf{S}_1(0) = \int_{-\infty}^{\infty} s_1(t)\, e^{-i\times 0\times t}\, dt = \int_{-\infty}^{\infty} s_1(t)\, dt \tag{2.69}$$

The right-hand side of this expression is none other than the area of the pulse $s_1(t)$. Thus, expression (2.68) may be written in the form

$$\mathbf{S}_1(\omega) = S_1(0)\,\frac{\sin(\omega\tau_p/2)}{\omega\tau_p/2} = S_1(0)\,\mathrm{sinc}\,(\omega\tau_p/2) \tag{2.68'}$$

Here $\mathrm{sinc}(\omega\tau_p/2)$ implies a function

$$\mathrm{sinc}\,(x) = (\sin x)/x$$

When lengthening (stretching) the pulse, the distance between the zeros of the function $S_1(\omega)$ is reduced, and this is equivalent to narrowing the spectrum: in this case the value of $S_1(0)$ increases. On the other hand, when shortening (compressing) the pulse, the distance between the zeros of the function $S_1(\omega)$ increases and this is equivalent to widening the spectrum; in this case the value of $S_1(0)$ decreases. As $\tau_p \to 0$ (A = const) the points $\omega_1 = \pm 2\pi/\tau_p$ corresponding to the first two zeros of the function $S_1(\omega)$ approach infinity, and the spectral density having an infinitely small magnitude becomes uniform in the frequency band from $-\infty$ to ∞.

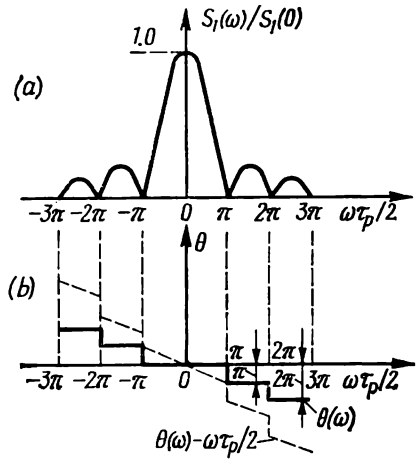


Fig. 2.15. (*a*) Modulus and (*b*) argument of the spectral density of a square pulse

Shown separately in Fig. 2.15 are the diagrams of the modulus $S_1(\omega)$ related to the value of $S_1(0)$ and of the argument $\theta(\omega)$ of the spectral density function. The first of these diagrams may be regarded as the amplitude characteristic, and the second, the phase characteristic of a square pulse spectrum.

Each change of the sign of $S_1(\omega)$ is taken into account in Fig. 2.15*b* by increasing the phase by π.

When the time is read not from the middle of the pulse (as in Fig. 2.13) but from its beginning (as in Fig. 2.16), the phase characteristic of the pulse spectrum must be supplemented with an added $\omega\tau_p/2$ which takes into account the $\tau_p/2$ delay shift of the pulse.

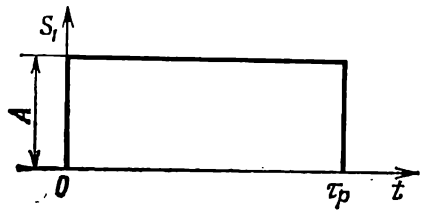


Fig. 2.16. Matching the time origin with the leading edge of a square pulse

In this case the resultant phase characteristic assumes a form shown in Fig. 2.15*b* by a broken line.

Let us consider the question of energy distribution in the spectrum of a pulse. According to Section 2.8 and formula (2.68′), the spectral

density of energy of a square pulse is

$$S^2(\omega) = S^2(0)\frac{\sin^2(\omega\tau_p/2)}{(\omega\tau_p/2)^2} = S^2(0)\left[\operatorname{sinc}\left(\frac{\omega\tau_p}{2}\right)\right]^2 \qquad (2.71)$$

Using Parseval's equation, one can easily determine the energy within a given frequency band.

Let us consider a band $\Delta\omega$ from $-\omega_1$ to ω_1. According to formula (2.66), the energy within this band is

$$E_{\Delta\omega} = \frac{1}{\pi}\int_0^{\omega_1} S^2(\omega)\,d\omega = A^2\tau_p^2\int_0^{\omega_1}\frac{\sin^2(\omega\tau_p/2)}{(\omega\tau_p/2)^2}\,d\omega$$

$$= A^2\tau_p^2\frac{1}{\pi}\frac{2}{\tau_p}\int_0^{\omega_1\tau_p/2}\frac{\sin^2 x}{x^2}\,dx = A^2\tau_p\eta\left(\frac{\omega_1\tau_p}{2}\right)$$

where $A^2\tau_p = E$ is the total energy of the pulse, while the function

$$\eta\left(\frac{\omega_1\tau_p}{2}\right) = \frac{2}{\pi}\int_0^{\omega_1\tau_p/2}\frac{\sin^2 x}{x^2}\,dx \qquad (2.72)$$

determines the relative amount of energy within the frequency band from 0 to ω_1.

The integral in expression (2.72) can be taken by parts and reduced to the form

$$\int_0^{\omega_1\tau_p/2}\frac{\sin^2 x}{x^2}\,dx = -\frac{\sin^2 x}{x}\Big|_0^{\omega_1\tau_p/2} + \int_0^{\omega_1\tau_p/2}\frac{2\sin x\cos x}{x}\,dx$$

$$= \frac{-\sin^2(\omega_1\tau_p/2)}{\omega_1\tau_p/2} + \int_0^{\omega_1\tau_p/2}\frac{\sin 2x}{x}\,dx = \frac{-2\sin^2(\omega_1\tau_p/2)}{\omega_1\tau_p} + \operatorname{si}\omega_1\tau_p$$

where $\operatorname{si}\, y = \int_0^y \frac{\sin x}{x}\,dx$, i.e. the sine integral.

Thus,

$$\eta\left(\frac{\omega_1\tau_p}{2}\right) = \frac{2}{\pi}\left[\operatorname{si}\omega_1\tau_p - \frac{2\sin^2(\omega_1\tau_p/2)}{\omega_1\tau_p}\right] \qquad (2.73)$$

The diagram of the function $\eta(\omega_1\tau_p/2)$ is shown in Fig. 2.17. It illustrates that when $\omega_1\tau_p/2 = \pi$ (i.e., when $f_1\tau_p = 1$), 90% of the total energy of the pulse is concentrated in the frequency

range from 0 to $f_1 = 1/\tau_p$. Formula (2.73) can be used for selecting the passband of a circuit (filter) for a given pulse energy utilization factor. It should be noted, however, that when it is necessary to

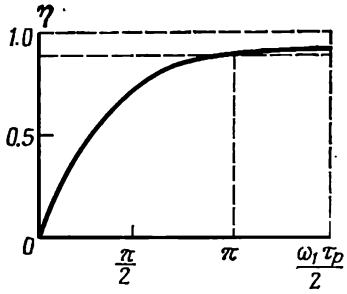


Fig. 2.17. Energy of a square pulse in the frequency band 0, ω_1

obtain a square-shaped pulse at the output of a filter, the magnitude of the product $f_1\tau_p$ must be much higher than unity.

2.9-2. Gaussian Pulse (Fig. 2.18*a*)

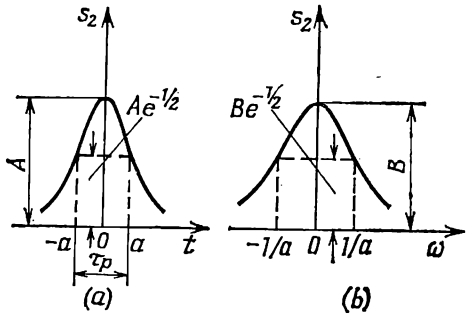


Fig. 2.18. (*a*) Gaussian pulse and (*b*) its spectral density

The pulse presented in Fig. 2.18*a* is determined by the expression

$$s_2(t) = Ae^{-t^2/2a^2} \quad (-\infty < t < \infty) \tag{2.74}$$

This pulse, which agrees in shape with the graph of the normal (Gaussian) distribution law is commonly referred to as a *Gaussian pulse*. The constant a is equal to a half of the pulse duration corresponding to the level $e^{-1/2} = 1/e^{1/2} = 0.606$ of the pulse amplitude. Thus, the total duration τ_p of the pulse is equal to $2a$.

Using expression (2.48), we obtain

$$S_2(\omega) = A\int_{-\infty}^{\infty} e^{-t^2/2a^2}e^{-i\omega t}\,dt \tag{2.75}$$

The integral is conveniently calculated by completing the square in the exponent in the integrand

$$-\left(\frac{t^2}{2a^2}+i\omega t\right) = -\left[\left(\frac{t^2}{2a^2}+i\omega t+d^2\right)-d^2\right] = -\left[\left(\frac{t}{\sqrt{2}a}+d\right)^2-d^2\right]$$

where the quantity d is determined from the condition

$$i\omega t = 2\left(t/\sqrt{2}\,a\right)d$$

whence

$$d = i\omega_a/\sqrt{2} \tag{2.76}$$

Thus, expression (2.75) can be written as

$$S_2(\omega) = Ae^{d^2}\int_{-\infty}^{\infty} e^{-(t/\sqrt{2}a+d)^2}\,dt$$

Taking a new variable $x = (t/\sqrt{2a}) + d$, we obtain

$$S_2(\omega) = Ae^{d^2}\sqrt{2}\,a\int_{-\infty}^{\infty} e^{-x^2}\,dx$$

Remembering that the integral in this expression is equal to $\sqrt{\pi}$, we may finally write

$$S_2(\omega) = A\sqrt{2\pi}\,ae^{-a^2\omega^2/2} = Be^{-\omega^2/2b^2} \tag{2.77}$$

where $b = 1/a$, $B = \sqrt{2\pi}\,aA$. The curve of this function is shown in Fig. 2.18*b*.

The result obtained is of great importance for the theory of signals. The Gaussian pulse and its spectrum turn out to be expressed by similar functions and are symmetrical: any of them can be obtained from the other by merely replacing t with ω or vice versa, as the case may be. The spectral band determined at a level of $e^{-1/2}$ of the maximum value is $2b = 2/a = 2^2/\tau_p = 4/\tau_p$, and the coefficient $B = \sqrt{2\pi}\,aA$.

The Gaussian spectrum

$$S_2(\omega) = Be^{-\omega^2/2b} \tag{2.78}$$

corresponds to the Gaussian pulse

$$s_2(t) = Ae^{-b^2t^2/2} = \frac{Bb}{\sqrt{2\pi}}e^{-b^2t^2/2} \tag{2.79}$$

with a duration of $2/b$ and an amplitude $A = Bb/\sqrt{2\pi}$.

It is clear that the shorter the duration τ_p of the pulse, the wider the spectral band $2b$.

Let us calculate the energy of the Gaussian pulse in the frequency band $\Delta\omega$ from $-\omega_1$ to ω_1. From formula (2.77) we obtain

$$E_{\Delta\omega} = \frac{1}{2\pi}\int_{-\omega_1}^{\omega_1}[A\sqrt{2\pi}\,ae^{-a^2\omega^2/2}]^2\,d\omega = A^2 2a^2\int_0^{\omega_1} e^{-a^2\omega^2}\,d\omega$$

$$= A^2 2a^2\frac{1}{a}\int_0^{a\omega_1} e^{-x^2}\,dx = \sqrt{\pi}\,A^2 a\frac{2}{\sqrt{\pi}}\int_0^{a\omega_1} e^{-x^2}\,dx$$

$$= \sqrt{\pi}\,A^2 a\Phi(a\omega_1) \tag{2.80}$$

Here $\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2}\,dx$ is the probability integral, and $\sqrt{\pi}A^2a = E$ is the total energy of the Gaussian pulse.

Thus, the ratio of the energy in the frequency band from $-\omega_1$ to ω_1 to the total energy of the Gaussian pulse is equal to $\Phi(a\omega_1)$.

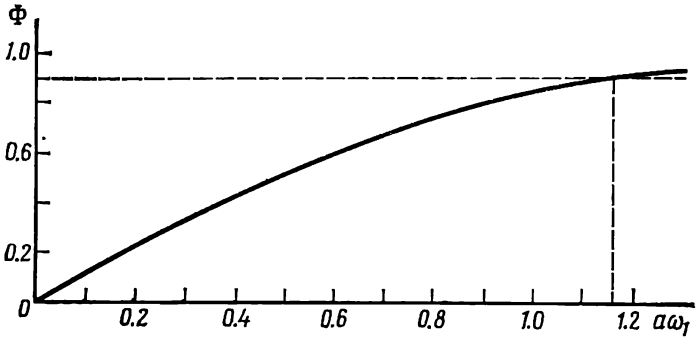


Fig. 2.19. Energy of a Gaussian pulse in the frequency band 0, ω_1

The function $\Phi(z)$ is tabulated; its curve is shown in Fig. 2.19. To obtain 90% of the pulse energy, it is necessary to satisfy the condition $z = a\omega_1 \approx 1.16$, or the product of the total duration $2a$

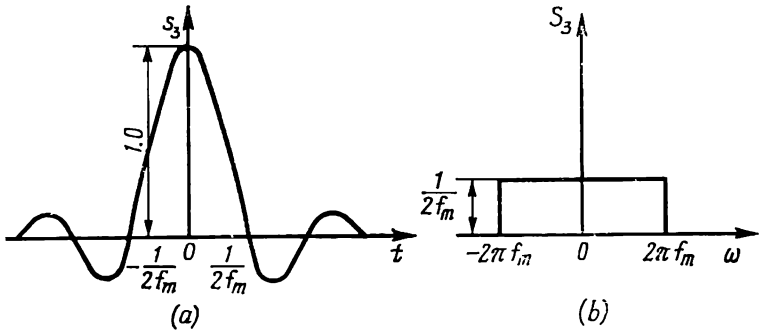


Fig. 2.20. (*a*) Pulse of the form sinc ($\omega_m t$) and (*b*) its spectral density

of the pulse and $2f_1$ should be

$$2a \times 2f_1 = z/\pi = 1.16/\pi \approx 0.4$$

2.9-3. Pulse of the sinc (x) Form

Shown in Fig. 2.20*a* is a pulse defined by the expression

$$s_3(t) = \operatorname{sinc}(\omega_m t) = \frac{\sin \omega_m t}{\omega_m t} = \frac{\sin 2\pi f_m t}{2\pi f_m t} \qquad (2.81)$$

Instead of calculating the spectral density by formula (2.48), let us use the results of Sec. 2.9-1 and the property of interchange-

ability of ω and t in the Fourier transforms for even functions of time (Sec. 2.7-7).

From Figs. 2.13 and 2.14 it is clear that, after replacing ω by t and t by ω, we will obtain a square-shaped spectrum corresponding to the given function $s_3(t)$. In order to use the transform (2.65), we have first to normalize the function $\mathbf{S}_1(\omega)$ in Fig. 2.14 in such a way as to make its maximum value $S_1(0)$ equal to unity (as $s_3(0) = s_3(t)$ in Fig. 2.20*a*). Putting $A\tau_p = 1$, we obtain the amplitude of the pulse $s_1(t)$ in Fig. 2.13 equal to $1/\tau_p$. Further, by replacing t by ω, $\tau_p/2$ by ω_m, and the pulse amplitude $1/\tau_p$ by $1/2\omega_m$, we obtain the spectral function

$$s_1(\omega) = \begin{cases} 1/2\omega_m & \text{if } |\omega| \leqslant \omega_m \\ 0 & \text{if } |\omega| > \omega_m \end{cases}$$

Using formula (2.65), we find the sought spectral density

$$\mathbf{S}_3(\omega) = 2\pi s_1(\omega) = \pi/\omega_m = 1/2f_m \quad \text{if} \quad |\omega| \leqslant \omega_m \tag{2.82}$$

The graph of the function $S_3(\omega)$ in shown in Fig. 2.20*b*.

2.9-4. Group of Identical and Equidistant Pulses

Let us denote the spectral density of the first pulse in a pulse train (Fig. 2.21) by $\mathbf{S}_1(\omega)$. Then, in view of formula (2.57), the spectral density of the second pulse delayed with respect to the

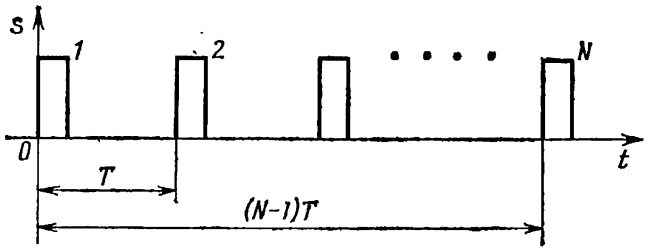


Fig. 2.21. Train of identical equidistant pulses

first one by a time T can be represented by an expression $\mathbf{S}_2(\omega) = \mathbf{S}_1(\omega)\, e^{-i\omega T}$, that of the third pulse, by $\mathbf{S}_3(\omega) = \mathbf{S}_1(\omega)\, e^{-i2\omega T}$, etc.

In accordance with the principle of linear summation of the spectra, a group of N pulses will have a spectral density

$$\mathbf{S}(\omega) = \mathbf{S}_1(\omega)\,[1 + e^{-i\omega T} + e^{-i2\omega T} + \ldots + e^{-i(N-1)\omega T}] \tag{2.83}$$

At frequencies satisfying the condition $\omega = k2\pi/T$, where k is an integer, each of the terms in the square brackets is equal to unity, therefore,

$$\mathbf{S}[k2\pi/T] = N\mathbf{S}_1[k2\pi/T] \tag{2.84}$$

Thus, at frequencies $\omega = k2\pi/T$ the modulus of the spectral density of the pulse train is N times greater than that of a single pulse. This is due to the fact that the spectral components of different pulses with the above mentioned frequencies are added together with phase shifts that are multiples of 2π.

At frequencies $\omega = (1/N)\ (2\pi/T)$, as well as at some other frequencies for which the sum of vectors e^{-ikT} vanishes, the total spec-

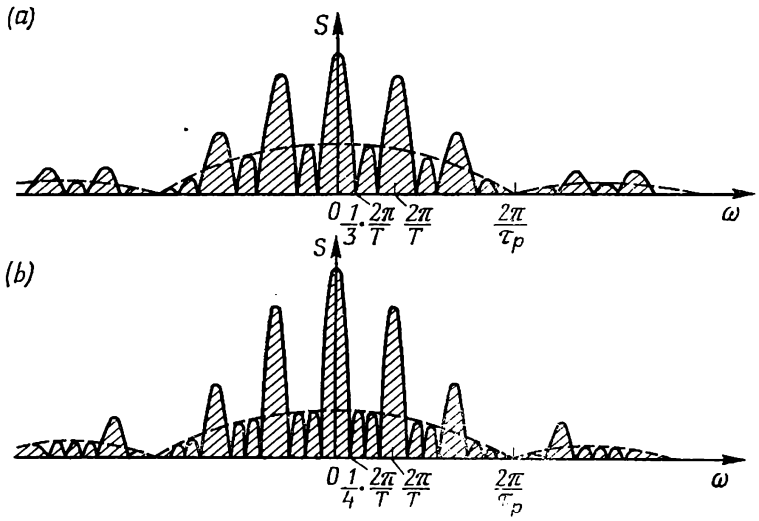


Fig. 2.22. Modulus of spectral density of a train of (*a*) three and (*b*) four pulses

tral density is zero. At intermediate frequencies the modulus $S(\omega)$ is determined as the vector sum of the spectral densities of individual pulses.

Figure 2.22*a* illustrates the spectral density (modulus) of a train of three square pulses, and Fig. 2.22*b*, that of a train of four pulses, the interval between the adjacent pulses being $T = 3\tau_p$. The broken lines indicate the spectral density of a single pulse. As the number of pulses in the train increases, the spectral density splits more and more and as $N \to \infty$ it assumes the line structure of the spectrum of a periodic function.

2.10. RELATION BETWEEN THE DURATION OF A SIGNAL AND THE WIDTH OF ITS SPECTRUM

From the preceding sections it is already clear that the shorter the duration of a signal, the wider its spectrum.

This fundamental proposition of the theory of signals can be established in the general form on the basis of the Fourier transform

$$\mathbf{S}(\omega) = \int_{-\infty}^{\infty} s(t)\, e^{-i\omega t}\, dt = \int_{-\infty}^{\infty} s(t) \cos \omega t\, dt - i \int_{-\infty}^{\infty} s(t) \sin \omega t\, dt \qquad (2.85)$$

Let us analyse the behaviour of each integral with an increase in ω. Riemann's lemma states that if a function $s(t)$ is absolutely integrable on a certain finite interval $[a, b]$, then

$$\lim_{\omega\to\infty}\int_a^b s(t)\cos\omega t\,dt = \lim_{\omega\to\infty}\int_a^b s(t)\sin\omega t\,dt = 0 \qquad (2.86)$$

The meaning of this statement is explained graphically in Fig. 2.23, whose upper portion (a) shows an arbitrary signal $s(t)$

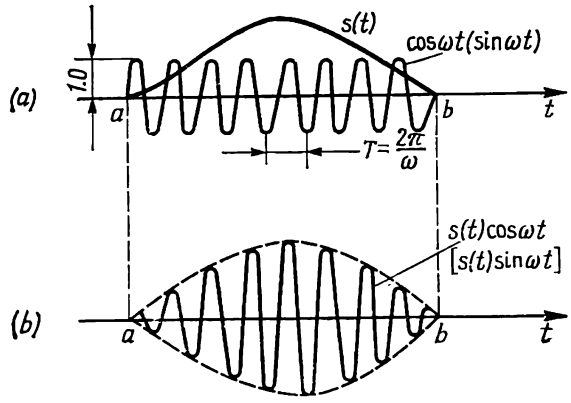


Fig. 2.23 Illustrating the relation between the duration and the spectrum width of a signal

and a harmonic oscillation with a frequency ω, and lower portion (b) illustrates the product $s(t)\cos\omega t$ [or $s(t)\sin\omega t$].

At a sufficiently high frequency ω each positive half wave in Fig. 2.23b is almost completely cancelled by the nearest negative half wave, and the total area under the curve $s(t)\cos\omega t$ [or $s(t)\times$ $\times\sin\omega t$] is close to zero. The "sufficiently high frequency" implies a frequency $\omega = 2\pi/T$ at which the period T is sufficiently small as compared with the duration of the signal $s(t)$.

It is evident that the shorter the signal, the smaller the period T satisfying this condition. In other words, the shorter the signal, the higher the boundary frequency of the spectrum. Since the lower boundary of the spectrum adjoins the zero frequency (in the case of signals without a high-frequency carrier as, for example, in Fig. 2.23a), the shorter the signal, the greater the total width of the spectrum. The product of the duration of a signal and the width of its spectrum cannot be smaller than a certain constant which is defined in Appendix 1.

2.11. INFINITELY SHORT PULSE OF UNIT AREA (DELTA FUNCTION)

Figure 2.24 illustrates some possible models of a pulse of unit area. The amplitudes of these pulses are inversely proportional to their widths (durations) defined in an appropriate manner. When the pulse duration tends to zero, the amplitude tends to infinity, while the pulse area remains constant and equal to unity.

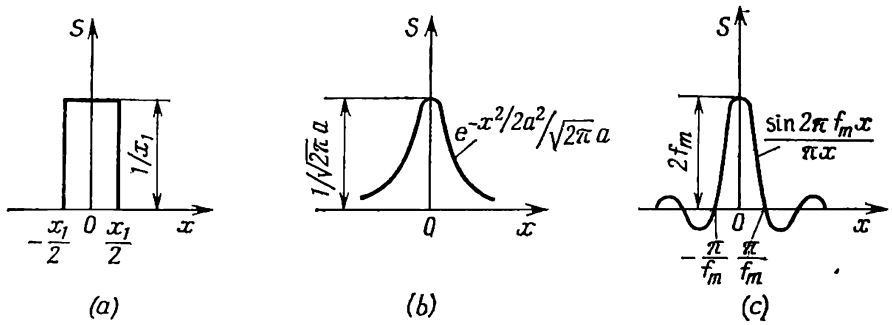


Fig. 2.24. Pulses transforming into the delta function as their duration approaches zero

In the case of a rectangular pulse (Fig. 2.24*a*), its amplitude must be taken equal to $1/x_1$, where x_1 is the pulse duration.

The amplitude of a Gaussian pulse (Fig. 2.24*b*) must be taken equal to $1/\sqrt{2\pi}a$, since

$$\int_{-\infty}^{\infty} e^{-x^2/2a^2}dx = \sqrt{2\pi}\,a$$

Finally, in the case of a $\sin(2\pi f_m x)/\pi x$ pulse of unit area (Fig. 2.24*c*) the amplitude is equal to $2f_m$ (when $x = 0$). The width of the pulse (the main lobe) is inversely proportional to the parameter f_m.

As the parameters x_1 and a tend to zero, and the parameter f_m tends to infinity, all three functions shown in Fig. 2.24 can be defined in the following manner:

$$\delta(x) = \begin{cases} \infty & \text{when } x = 0 \\ 0 & \text{when } x \neq 0 \end{cases} \tag{2.87}$$

while at the same time

$$\int_{-\infty}^{\infty} \delta(x)\,dx = \text{pulse area} = 1 \tag{2.88}$$

The function $\delta(x)$ having the above-said properties is called *unit impulse*, *unit-impulse function*, *delta function* or *Dirac delta function*.

As applied to the functions shown in Fig. 2.24*b* and *c*, the delta function must be defined by the expressions

$$\delta(x) = \lim_{a \to 0} \frac{1}{\sqrt{2\pi}\, a} e^{-x^2/2a^2} \tag{2.89}$$

$$\delta(x) = \lim_{f_m \to \infty} [\sin(2\pi f_m x)/\pi x] \tag{2.90}$$

Other numerous definitions of the function $\delta(x)$ are also possible.

For a pulse shifted along the x-axis by a value x_0, definitions (2.87) through (2.90) should be written in a more general form

$$\delta(x - x_0) = \begin{cases} \infty & \text{when } x = x_0 \\ 0 & \text{when } x \neq x_0 \end{cases} \tag{2.87'}$$

$$\int_{-\infty}^{\infty} \delta(x - x_0)\, dx = 1 \tag{2.88'}$$

$$\delta(x - x_0) = \lim_{a \to 0} \frac{1}{\sqrt{2\pi}\, a} e^{-(x - x_0)/2a^2} \tag{2.89'}$$

$$\delta(x - x_0) = \lim_{f_m \to \infty} \frac{\sin 2\pi f_m (x - x_0)}{\pi (x - x_0)} \tag{2.90'}$$

The function $\delta(x)$ has some important properties and is widely used in mathematics, physics, and engineering.

Definitions (2.87') and (2.88') yield the following basic relationships

$$\int_{-\infty}^{\infty} \delta(x - x_0) f(x)\, dx = f(x_0) \int_{-\infty}^{\infty} \delta(x - x_0)\, dx = f(x_0) \tag{2.91}$$

Since the function $\delta(x - x_0)$ is equal to zero along the entire x-axis, except for the point $x = x_0$ (where it is infinitely large), the interval of integration can be made as small as desired so long as it contains the point x_0. Within this interval the function $f(x)$ has a constant value $f(x_0)$ which can be taken outside the integral sign. So, the multiplication of any integrand $f(x)$ by $\delta(x - x_0)$ makes it possible to equate the integral of the product to the value of $f(x)$ at the point $x = x_0$.

The relation (2.91) is known in mathematics as the "*filtering property*"* of the delta function.

The theory of signals deals with delta functions of the arguments t or ω, depending on the domain in which the function is investigated, i.e., time or frequency.

First, let us consider the properties of the function $\delta(t)$. In this case the spectral characteristic of the delta function is of primary

* From the viewpoint of radio engineering the term "gating" is more appropriate.

importance. In Sec. 2.9 it was found that as the duration τ_p of a square pulse (having a constant amplitude) decreases, the width of the major lobe of the spectral density function increases, while its value $S(0)$ diminishes quickly. In our case, however, when a decrease in the pulse duration is accompanied by an increase in its amplitude, the value of the spectral density remains constant and equal to $S(0) = 1$ for all frequencies in the interval $-\infty < \omega < \infty$. The same phenomenon takes place when any pulse of unit area is shortened.

The spectral density of a delta function is therefore a real number and is equal to unity for all frequencies. From this fact it follows that the phase characteristic of a delta function $\delta(t)$ is equal to zero for all frequencies. This means that the sum of all the harmonics of a unit impulse, having zero epoch angles, has a peak that is infinitely large at $t = 0$.

Likewise, the function $\delta(t - t_0)$ defining a unit impulse at the instant t_0 has a spectral density $\mathbf{S}(\omega) = e^{-i\omega t_0}$. The modulus of this function is still equal to unity, while the phase characteristic $\theta(\omega) = \omega t_0$.

The spectral density of the delta function can also be obtained in a formal way, by using the Fourier transform:

$$\mathbf{S}(\omega) = \int_{-\infty}^{\infty} \delta(t - t_0)\, e^{-i\omega t}\, dt$$

Using expression (2.91), we obtain

$$\mathbf{S}(\omega) = e^{-i\omega t_0} \int_{-\infty}^{\infty} \delta(t - t_0)\, dt = e^{-i\omega t_0} \tag{2.92}$$

In the particular case $t_0 = 0$ we get $\mathbf{S}(\omega) = 1$.

Likewise, $\delta(t - t_0)$ can be written in the form of the inverse Fourier transform of $\mathbf{S}(\omega) = e^{-i\omega t_0}$:

$$\delta(t - t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, e^{-i\omega t}\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(t - t_0)}\, d\omega \tag{2.93}$$

The energy of a unit impulse is infinitely large. In the case of a spectral approach, this follows from Parseval's equation [see (2.66)] which yields infinity for $S(\omega) = 1$, while in the case of a time approach, this follows from the fact that the energy of a pulse, which is proportional to the square of the pulse amplitude (i.e., to the quantity $1/\tau_p^2$) and the first power of the pulse duration τ_p, rises as $1/\tau_p$ as the pulse is shortened. As $\tau_p \to 0$ the energy becomes infinitely large.

The concept of the unit impulse is widely used in the study of the action of short pulses on linear systems. The amplitude of a real pul-

se need not be infinitely large and the duration of the pulse, infinitely small. It is quite sufficient that the duration of the pulse is short as compared to the time constant of the system in question (or as compared to the period of the natural oscillation of the system).

Now let us investigate the properties of $\delta(\omega)$. Whatever has been said of the properties of $\delta(t)$ can be applied to $\delta(\omega)$ on replacing t with ω and ω with t.

By analogy with expression (2.93), we can write

$$\delta(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t}\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t}\, dt \qquad (2.94)$$

In this case the reversal of the sign in the exponent does not affect the magnitude of the integral (see Sec. 2.7-7*a*).

Correspondingly,

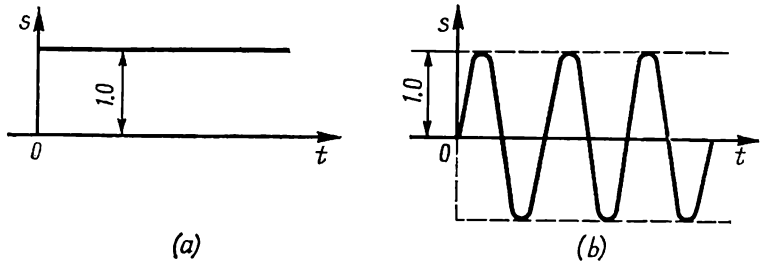
$$\delta(\omega - \omega_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\omega-\omega_0)t}\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i(\omega-\omega_0)t}\, dt \qquad (2.94')$$

2.12. SPECTRA OF SOME NONINTEGRABLE FUNCTIONS

One of the conditions for the applicability of the Fourier transform to a function is its absolute integrability:

$$\int_{-\infty}^{\infty} |f(t)|\, dt < \infty \qquad (2.95)$$

This condition substantially limits the class of functions for which there exists the Fourier spectrum expressed by ordinary functions. For example, such important functions as a unit step (as in



(*a*) (*b*)

Fig. 2.25. Examples of functions that do not satisfy the condition of absolute integrability

Fig. 2.25*a*) or a harmonic oscillation (as in Fig. 2.25*b*) initiated at some moment, do not satisfy the condition (2.95). This difficulty can be obviated by generalizing the Fourier transform in such a way as to ensure the integrability of a certain auxiliary function.

The method utilizing the so-called "convergence factor" has been widely used in mathematics for a long period of time. According to this method, the unit step is first replaced by the exponential pulse e^{-ct} $(c > 0)$ for which the condition (2.95) is satisfied and the spectral density can easily be found:

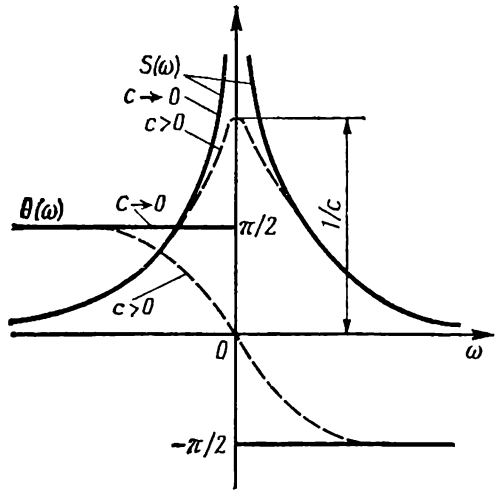


Fig. 2.26. Modulus and argument o he spectral density of an exponential pulse

$$S(\omega) = \int_0^\infty e^{-ct} e^{-i\omega t}\, dt$$

$$= -\frac{1}{c+i\omega} e^{-(c+i\omega)t} \Big|_{t=0}^{\infty}$$

$$= \frac{1}{c+i\omega} \qquad (2.96)$$

By expressing $S(\omega)$ in the form of (2.50), we obtain

$$S(\omega) = S(\omega)\, e^{i\theta(\omega)} = \frac{1}{\sqrt{c^2+\omega^2}} e^{-i \arctan \omega/c}, \quad (-\infty < \omega < \infty) \qquad (2.96')$$

The graphs of the modulus $S(\omega) = 1/\sqrt{c^2 + \omega^2}$ (amplitude-frequency characteristic) and the argument $\theta(\omega) = -\arctan(\omega/c)$ (phase-frequency characteristic) of the spectral density of the exponential pulse $s(t) = e^{-ct}$ are shown in Fig. 2.26 by broken lines.

Letting c approach zero, we obtain in the limit the following expressions for the spectral density of the unit step:

$$S(\omega) = \lim_{c\to 0} \frac{1}{c+i\omega} = \frac{1}{i\omega}$$

$$= \begin{cases} \dfrac{1}{\omega} e^{-i\pi/2}, & \theta(\omega) = -\pi/2 \quad \text{when } \omega > 0 \\ \dfrac{1}{|\omega|} e^{i\pi/2}, & \theta(\omega) = +\pi/2 \quad \text{when } \omega < 0 \end{cases} \qquad (2.97)$$

The graphs of $S(\omega)$ and $\theta(\omega)$ calculated from these formulas are shown in Fig. 2.26 by solid lines. It should be remembered, however, that in some cases formula (2.97) may lead to misunderstanding. To obtain a correct result through the use of the convergence factor, the spectral density calculated for $c \neq 0$ must be substituted into the Fourier integral, while the passage to the limit $c \to 0$ must be performed only in the final result, after the Fourier integral has been calculated. This procedure is similar to the transform from the

variable ω to the complex variable $p = \sigma + i\omega$ in conjunction with the appropriate selection of the integration path in the p plane. Such a method leading to the Laplace transformation will be used in Sec. 2.13. But in cases where the Fourier spectrum itself must be used the convergence factor e^{-ct} need not be introduced. The use of the so-called "generalized functions", including the delta function, would be more effective.

The properties of the delta function discussed in the preceding section permit the concept of spectral density to be extended to cover a harmonic or, in general, any periodic oscillation.

As an illustration, let us take a harmonic oscillation $s(t) = A_0 \cos(\omega_0 t + \theta_0)$ and, ignoring the fact that such a signal is not absolutely integrable, write the expression for the spectral density in the form (2.48):

$$S(\omega) = \int_{-\infty}^{\infty} s(t)\, e^{-i\omega t}\, dt = A_0 \int_{-\infty}^{\infty} \cos(\omega_0 t + \theta_0)\, e^{-i\omega t}\, dt$$

$$= \frac{A_0 e^{i\theta_0}}{2} \int_{-\infty}^{\infty} e^{-i(\omega-\omega_0)t}\, dt + \frac{A_0 e^{-i\theta_0}}{2} \int_{-\infty}^{\infty} e^{-i(\omega+\omega_0)t}\, dt$$

From formula (2.94′) we obtain

$$S(\omega) = \frac{A_0}{2} [2\pi e^{i\theta_0} \delta(\omega - \omega_0) + 2\pi e^{-i\theta_0} \delta(\omega + \omega_0)]$$

$$= A_0 \pi [e^{i\theta_0} \delta(\omega - \omega_0) + e^{-i\theta_0} \delta(\omega + \omega_0)] \qquad (2.98)$$

This function is zero for all frequencies except for $\omega = \omega_0$ and $\omega = -\omega_0$ at which $S(\omega)$ becomes infinite. As could be expected, a harmonic oscillation with a *finite* amplitude corresponds to an *infinitely large* spectral density at discrete frequencies ω_0 and $-\omega_0$.

In particular, while equating ω_0 to zero, we obtain the spectral density of a signal in the form of a d-c voltage A_0:

$$S(\omega) = A_0 2\pi \delta(\omega) \qquad (2.99)$$

By applying expression (2.98) to all the harmonics of any periodic signal

$$s(t) = A_0 + \sum_{n=1}^{\infty} A_n \cos(n\omega_1 t + \theta_n)$$

we can introduce the concept of the spectral density of a periodic signal in the form of a sum of delta functions

$$S(\omega) = A_0 2\pi \delta(\omega) + A_1 \pi [e^{i\theta_1} \delta(\omega - \omega_1) + e^{-i\theta_1} \delta(\omega + \omega_1)]$$
$$+ A_2 \pi [e^{i\theta_2} \delta(\omega - 2\omega_1) + e^{-i\theta_2} \delta(\omega + 2\omega_1)]$$
$$+ \ldots + A_n \pi [e^{i\theta_n} \delta(\omega - n\omega_1) + e^{-i\theta_n} \delta(\omega + n\omega_1)] + \ldots \qquad (2.100)$$

Such an approach turns out to be very useful when studying a combination of a pulse signal and monochromatic oscillations.

For instance, let us find the spectrum of the sum of two signals: a pulse signal $s_1(t)$ and a monochromatic signal $s_2(t) = A_0 \cos \omega_0 t$ (see Fig. 2.27*a*). By applying expression (2.48) to $s_1(t)$, we obtain the conventional spectral density $S_1(\omega)$ defining a continuous spectrum (the dashed area in Fig. 2.27*b*). On the other hand, applying (2.48) to $s_2(t)$, we obtain the spectrum defined by expression (2.98).

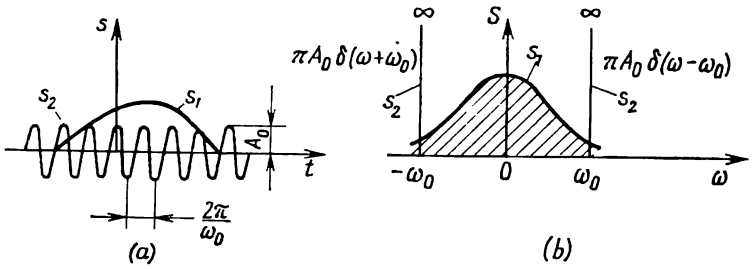


Fig. 2.27. (*a*) Pulse and monochromatic signals and (*b*) their spectral densities

In Fig. 2.27*b* this spectrum is illustrated by two spectral lines extending to infinity.

Now let us find the spectrum of a unit step. Let this function be represented by a rectangular pulse with the beginning at the point $t = 0$ and the end at the point T tending to infinity (see Fig. 2.28*a*). For such a representation, the spectral density of the unit step can be determined by the expression

$$S(\omega) = \lim_{T\to\infty} \int_0^T e^{-i\omega t}\, dt = \lim_{T\to\infty} \left\{ \frac{1}{-i\omega} e^{-i\omega t} \Big|_0^T \right\}$$

$$= \lim_{T\to\infty} \left\{ \frac{1 - e^{-i\omega T}}{i\omega} \right\} = \lim_{T\to\infty} \left\{ \frac{\sin \omega T}{\omega} + \frac{1 - \cos \omega T}{i\omega} \right\} \qquad (2.101)$$

But from formula (2.90)

$$\lim_{T\to\infty} \frac{\sin \omega T}{\omega} = \pi\delta(\omega) \qquad (2.102)$$

Consequently, the spectrum of the unit step is

$$S(\omega) = \pi\delta(\omega) + (1 - \lim_{T\to\infty} \cos \omega T)/i\omega \qquad (2.103)$$

The first term on the right-hand side of this expression determines the spectrum of a d-c voltage $A_0 = 1/2$ [see (2.99)] shown in Fig. 2.28*b*, while the second term determines the spectrum of the

function shown in Fig. 2.28*c*. The sum of these two functions forms the unit step at the moment $t = 0$ (see Fig. 2.28*a*).

Now let us consider the properties of the function

$$\varkappa(\omega) = \lim_{T\to\infty} \cos \omega T \tag{2.104}$$

For any nonzero magnitudes of ω this function is indeterminate and can take on any value within the range $(-1, +1)$, whereas

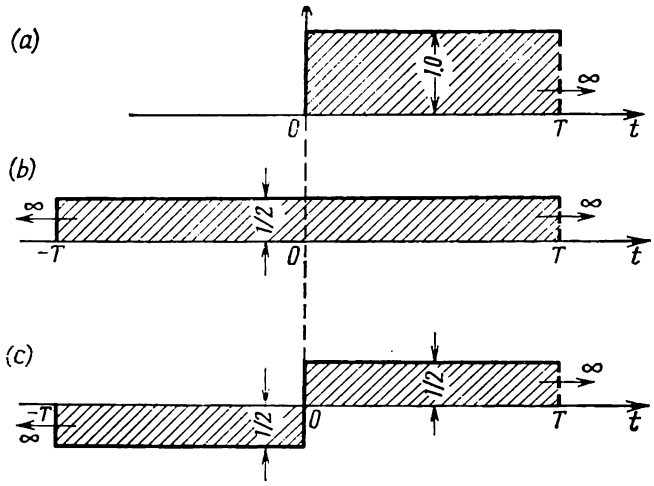


Fig. 2.28. Defining the spectrum of a unit step

at the point $\omega = 0$ it has a definite value that can readily be found. Since we have adopted that the passage to the limit is performed at the final stage, we find that for $\omega = 0$ $\varkappa(0) = 1$ irrespective of the value of T.

Another evident property of $\varkappa(\omega)$ is that the integral of $\varkappa(\omega)$ taken over any finite interval (a, b) is equal to zero

$$\int_a^b \varkappa(\omega)\, d\omega = \lim_{T\to\infty} \int_a^b \cos \omega T\, d\omega = \lim_{T\to\infty} \frac{\sin \omega T \Big|_a^b}{T} = 0$$

Moreover, in view of the Riemann lemma discussed in Sec. 2.10, for any function $g(\omega)$ that is absolutely integrable in the finite interval (a, b), the following identities hold:

$$\lim_{T\to\infty} \int_a^b g(\omega) \cos \omega T\, d\omega \equiv \lim_{T\to\infty} \int_a^b g(\omega) \sin \omega T\, d\omega \equiv 0$$

By analogy with the procedure adopted in the theory of generalized functions, we may assume that a rapidly oscillating function $\varkappa(\omega)$ is zero for all $\omega \neq 0$.

Thus, we may introduce the following definition:

$$\varkappa(\omega) = \lim_{T\to\infty} \cos \omega T = [1(\omega)] = \begin{cases} 1 \text{ when } \omega = 0 \\ 0 \text{ when } \omega \neq 0 \end{cases} \quad (2.105)$$

The function $\varkappa(\omega) = [1(\omega)]$ thus defined may be called a "*needle*" function. A graphic representation of the "needle" function is shown in Fig. 2.29.

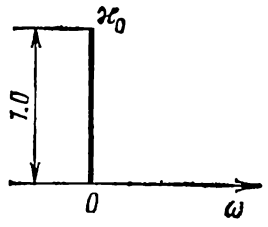


Fig. 2.29. Needle function

Applying the above definitions, formula (2.103) may be written in the form:

$$S(\omega) = \pi\delta(\omega) + \frac{1-[1(\omega)]}{i\omega} \quad (2.106)$$

From this expression it is evident that when $\omega = 0$ the second term is zero (since $[1(\omega)] = 1$ when $\omega = 0$) and the entire contribution to the spectrum comes from the term $\pi\delta(\omega)$ corresponding to the constant component 1/2, while at nonzero frequencies ω the first term is zero and the second one is equal to $1/i\omega$ (since $[1(\omega)] = 0$ when $\omega \neq 0$) [9].

From this it follows that when studying the action of a unit step on circuits whose transfer function is zero when $\omega = 0$ (i.e., circuits which do not let pass through direct current), the spectral density of a unit step can be represented in the form of expression (2.97).

2.13. REPRESENTATION OF SIGNALS ON THE PLANE OF A COMPLEX VARIABLE

The analysis of transmission of signals through linear circuits described by a complex transfer function is much facilitated when using the methods of contour integration on the plane of complex frequency $p = \sigma + i\omega$. Also, the transform from the real variable ω to $p = \sigma + i\omega$ makes it possible to completely eliminate the restrictions stemming from the requirement of the absolute integrability of the function $s(t)$.

Let us assume that the function $s(t)$, which generally exists for $-\infty < t < \infty$, is a sum of two functions

$$s(t) = s_+(t) + s_-(t),$$

of which $s_+(t)$ is defined for $0 < t < \infty$, while $s_-(t)$ is defined for $-\infty < t < 0$.

Using the pair of Fourier transforms (2.48), (2.49), let us proceed from ω to p at first for the function $s_+(t)$. For this purpose, let

$s_+(t)$ be multiplied by $e^{-\sigma_1 t}$, where $\sigma_1 > 0$ is selected so as to provide for the absolute integrability of the function $e^{-\sigma_1 t} s_+(t)$ over the interval $0 < t < \infty$. After substituting $p = \sigma_1 + i\omega$, expression (2.49) can be reduced to

$$e^{-\sigma_1 t} s_+(t) = \frac{1}{2\pi i} \int_{\sigma_1 - i\infty}^{\sigma_1 + i\infty} S_+\left(\frac{p-\sigma_1}{i}\right) e^{(p-\sigma_1)t}\, dp$$

or

$$s_+(t) = \frac{1}{2\pi i} \int_{\sigma_1 - i\infty}^{\sigma_1 + i\infty} L_{s+}(p)\, e^{pt}\, dp \qquad (2.107)$$

where

$$L_{s+}(p) = S_+\left(\frac{p-\sigma_1}{i}\right) = \int_0^\infty s_+(t)\, e^{-\sigma_1 t} e^{-i\omega t}\, dt = \int_0^\infty s_+(t)\, e^{-pt}\, dt \qquad (2.108)$$

which is called the *Laplace transform* of the function $s_+(t)$.

Relation (2.107), by analogy with expression (2.49), is often called *inverse Laplace transform.*

From a comparison of expressions (2.107) and (2.49) it is clear that the transform from ω to p means the change of the integration

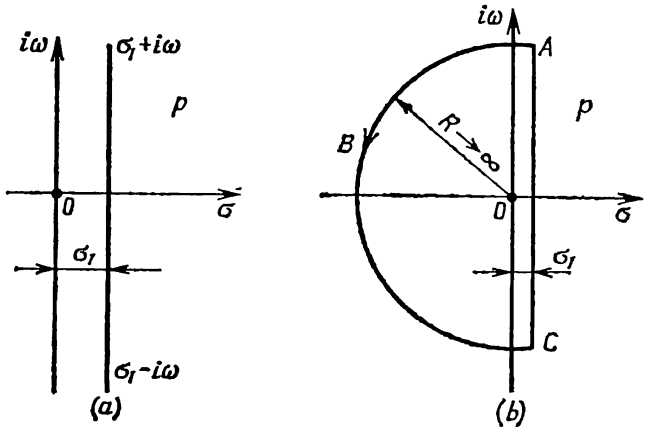


Fig. 2.30. (*a*) Integration path along the straight line $\sigma_1 - i\infty$, $\sigma_1 + i\infty$ on the p-plane and (*b*) formation of a closed contour by adding the arc ABC with $R \to \infty$

path. In expression (2.49) the integration is along the real axis ω, while in expression (2.107) it is along a straight line extending parallel to the imaginary axis $i\omega$ at a distance of σ_1 to the right of this axis (Fig. 2.30*a*). The constant σ_1 is determined by the character of the integrand in (2.107); the integration path must pass to the right of the poles of this function.

By adding an arc of an infinitely large radius to the straight line $\sigma_1 - i\infty$, $\sigma_1 + i\infty$, we can form a closed integration contour (Fig. 2.30*b*). To avoid any changes in the value of the integral due to the addition of this arc, one should observe the following rule: with positive values of t the contour must be located in the left-hand half plane of the variable p, and with negative t it must be located in the right-hand half plane.

Then, in the first case $t > 0$ [when the arc is described in the left-hand half plane (Fig. 2.31*a*)] the integration contour encompasses

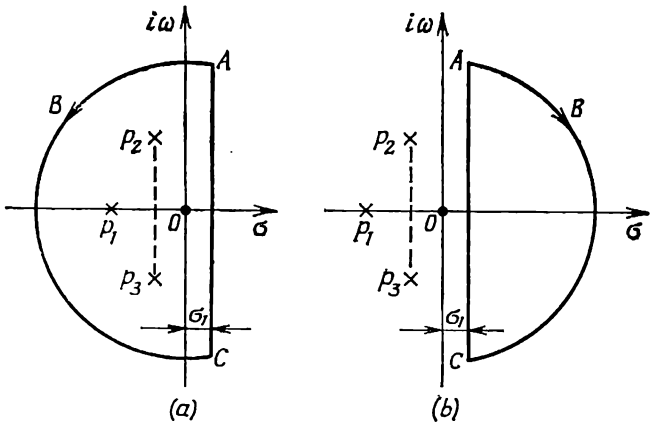


Fig. 2.31. Closing the integration contour to represent the function $s_+(t)$ for (*a*) $t > 0$ and (*b*) $t < 0$

all poles of the integrand (those lying to the left of the straight line $\sigma_1 - i\infty$, $\sigma_1 + i\infty$) and, in accordance with the residue theorem, the integral (2.107) is defined as

$$s_+(t) = \frac{1}{2\pi i}\int\limits_{\sigma_1 - i\infty}^{\sigma_1 + i\infty} L_{s+}(p)\, e^{pt}\, dp = \frac{1}{2\pi i}\oint\limits_{ABCA} L_{s+}(p)\, e^{pt}\, dp = \sum \text{res} \tag{2.109}$$

where $\sum$ res is the sum of the residues at the poles of the integrand.

When a similar arc is described in the right-hand half plane, i.e., with $t < 0$ (Fig. 2.31*b*), the poles of the function $L_{s+}(p)\, e^{pt}$ are beyond the integration contour and, in accordance with the Cauchy theorem, the integral along the closed contour is zero.

Thus, depending on the method of closing the integration contour, we obtain

for $t > 0$ (contour in Fig. 2.31*a*)

$$s_+(t) = \frac{1}{2\pi i}\int\limits_{\sigma_1 - i\infty}^{\sigma_1 + i\infty} L_{s+}(p)\, e^{pt}\, dp = \frac{1}{2\pi i}\oint\limits_{ABCA} L_{s+}(p)\, e^{pt}\, dp = \sum \text{res} \tag{2.110}$$

for $t < 0$ (contour in Fig. 2.31*b*)

$$s_+(t) = \frac{1}{2\pi i}\int_{\sigma_1-i\infty}^{\sigma_1+i\infty} L_{s+}(p)\, e^{pt}\, dp = \frac{1}{2\pi i}\oint_{ABCA} L_{s+}(p)\, e^{pt}\, dp = 0 \quad (2.111)$$

Let us recall an important property of contour integrals: the magnitude of the integral is independent of the shape of the closed contour along which it is taken, provided that the poles of the integrand remain within the contour. On the basis of this property the contour formed by adding the arc ABC of an infinitely large radius (Fig. 2.31*a*) to the straight line $\sigma_1 - i\infty$, $\sigma_1 + i\infty$ can be arbitrarily deformed provided that all the poles located to the left of the line remain inside the contour.

Thus, the calculation of the integral (2.109) is reduced to the determination of the residues at the poles of the integrand.

A similar reasoning may be applied to the function $s_-(t)$ defined for $-\infty < t < 0$. By multiplying $s_-(t)$ by $e^{-\sigma_2 t}$, with $\sigma_2 < 0$ selected so as to provide for the absolute integrability of the function $e^{-\sigma_2 t} s_-(t)$ over the interval $-\infty < t < 0$, we may write

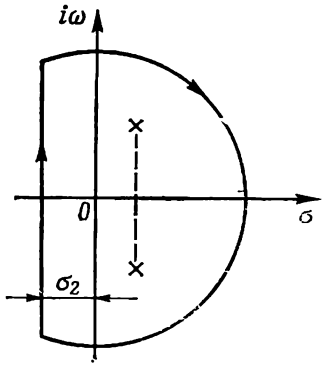


Fig. 2.32. Closing the integration contour to represent the function $s_-(t)$ for $t < 0$

$$L_{s-}(p) = \int_{-\infty}^{0} s_-(t)\, e^{-\sigma_2 t} e^{-i\omega t}\, dt = \int_{-\infty}^{0} s_-(t)\, e^{-pt}\, dt \quad (2.112)$$

$$s_-(t) = \frac{1}{2\pi i}\int_{\sigma_2-i\infty}^{\sigma_2+i\infty} L_{s-}(p)\, e^{pt}\, dp \quad (2.113)$$

The integration contour for this case is shown in Fig. 2.32. The integral is equal to the sum of the residues at the poles of the function $L_{s-}(p)\, e^{pt}$, located in the right-hand half plane p. This sum should be taken with the minus sign, since with $t < 0$ the contour is followed clockwise.

Expressions (2.108), (2.112) and (2.107), (2.113) can be combined as follows

$$L_s(p) = L_{s+}(p) + L_{s-}(p) \quad (2.114)$$

$$s(t) = \frac{1}{2\pi i}\left[\int_{\sigma_1-i\infty}^{\sigma_1+i\infty} L_{s+}(p)\, e^{pt}\, dp + \int_{\sigma_2-i\infty}^{\sigma_2+i\infty} L_{s-}(p)\, e^{pt}\, dp\right] \quad (2.115)$$

The relation (2.114) is called *bilateral Laplace transform.*

Shown in Fig. 2.33 are the regions of convergence of the functions $L_{s+}(p)$ and $L_{s-}(p)$ on the plane p. For $L_{s+}(p)$ this region is located to the right of the straight line $\sigma = -\sigma_1$ on which the complex-conjugate poles are disposed, while for $L_{s-}(p)$ it is located to the

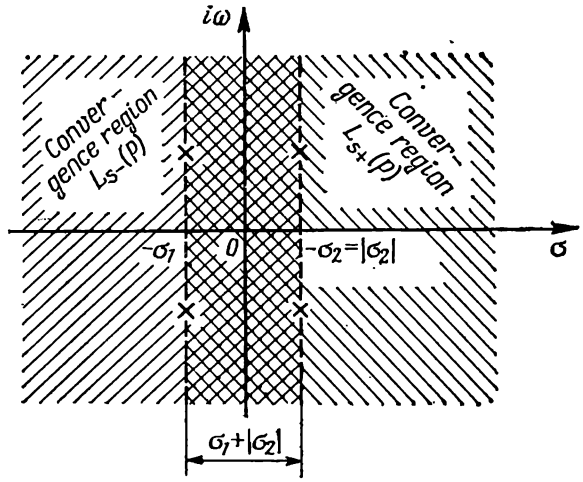


Fig. 2.33. Regions of convergence for the bilateral Laplace transform

left of the straight line $\sigma = |\sigma_2|$. The region of convergence for $L_s(p)$ is a strip having a width $\sigma_1 + |\sigma_2|$ (Fig. 2.33).

The path of integration must pass along a straight line located inside this strip and parallel to the axis $i\omega$ and then along a closing arc disposed in the left-hand half plane for $t > 0$ and, respectively, in the right-hand half plane for $t < 0$.

The one-sided Laplace transformation has gained wide application in the analysis of transient processes associated with the action of an external force on a circuit, when the time origin is made coincident with the beginning of the action. The two-sided (bilateral) Laplace transformation finds an ever greater application in the analysis of processes and functions which are inherently bilateral (e.g., correlation functions discussed in Sec. 2.16).

In the analysis of even functions $s(t) = s(-t)$, when it may be assumed that $s_+(t) = s_-(-t)$, the following relation holds:

$$L_{s-}(p) = \int_{-\infty}^{0} s_-(t)\, e^{-pt}\, dt = \int_{\infty}^{0} s_-(-t)\, e^{-p(-t)}\, dt$$

$$= \int_{0}^{\infty} s_+(t)\, e^{pt}\, dt = L_{s+}(-p) \qquad (2.116)$$

Let us illustrate the use of expressions (2.114) through (2.116) with an example of a function of time (Fig. 2.34) with $\alpha_1 > 0$ and

$\alpha_2 < 0$:

$$s(t) = \begin{cases} e^{|\alpha_2| t} & \text{when } t \leqslant 0 \\ e^{-\alpha_1 t} & \text{with } t \geqslant 0 \end{cases}$$

Using formulas (2.112) and (2.108), we find

$$L_{-s}(p) = \int_{-\infty}^{0} e^{-\alpha_2 t} e^{-pt}\, dt = -\frac{1}{\alpha_2 + p}$$

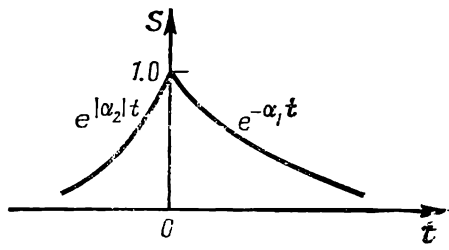


Fig. 2.34. Example of a time function requiring the application of the bilateral Laplace transform

the pole $p_n = -\alpha_2 = |\alpha_2|$:

$$L_{s+}(p) = \int_{0}^{\infty} e^{-\alpha_1 t} e^{-pt}\, dt = \frac{1}{\alpha_1 + p}$$

the pole $p_n = -\alpha_1$:

$$L_s(p) = -\frac{1}{\alpha_2 + p} + \frac{1}{\alpha_1 + p} = \frac{|\alpha_2| - \alpha_1}{(p + \alpha_1)(p + \alpha_2)}$$

$$= -\frac{|\alpha_2| + \alpha_1}{p^2 + (\alpha_1 - |\alpha_2|)\, p - \alpha_1 |\alpha_2|} \qquad (2.117)$$

In the particular case $|\alpha_2| = \alpha_1$ we obtain

$$L_{s-}(p) = L_{s+}(-p) = \frac{1}{\alpha_1 - p}, \quad L_s(p) = \frac{1}{\alpha_1 - p} + \frac{1}{\alpha_1 + p} = \frac{2\alpha_1}{\alpha_1^2 + p^2} \qquad (2.118)$$

It should be noted that to transform from the Laplace transform to the Fourier spectrum S (ω), it is sufficient to replace p by $i\omega$ in expression (2.118).

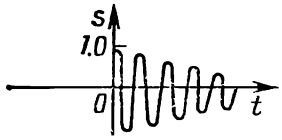
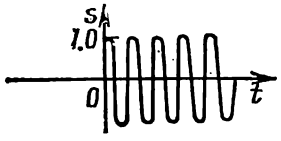
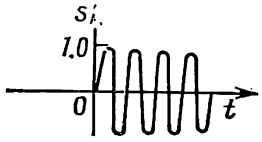
Thus, in the given case we have

$$S(\omega) = 2\alpha_1 / (\alpha - \omega^2) \qquad (2.119)$$

Table 2.1 gives the Laplace transforms and the corresponding Fourier spectra of some functions that are widely used in the theory of signals.

Definition of function $s(t)$
$\delta(t)$
$s(t)=\begin{cases}1 \text{ when } t\geqslant 0\\ 0 \text{ when } t<0\end{cases}$
$s(t)=\begin{cases}e^{-\alpha t} \text{ when } t\geqslant 0,\\ 0 \text{ when } t<0\end{cases}$ $\alpha>0$
$s(t)=\begin{cases}e^{\lvert\alpha\rvert t} \text{ when } t\leqslant 0,\\ 0 \text{ when } t>0\end{cases}$ $\alpha<0$
$s(t)=\begin{cases}e^{\lvert\alpha_2\rvert t} \text{ when } t\leqslant 0,\\ e^{-\alpha_1 t} \text{ when } t\geqslant 0\end{cases}$ $\alpha_2<0,\ \alpha_1>0$
$s(t)=\begin{cases}e^{-\alpha t}\cos\omega_0 t \text{ when } t\geqslant 0,\\ 0 \text{ when } t<0\end{cases}$ $\alpha>0$

Laplace transform	Spectral density $S(i\omega)$
1	1
$\frac{1}{p}$	$\pi\delta(\omega)+\frac{1-[1(\omega)]}{i\omega}$ (see Sec. 2.12)
$\frac{1}{\alpha+p}$	$\frac{1}{\alpha+i\omega}$
$-\frac{1}{\alpha+p}$	$-\frac{1}{\alpha+i\omega}$
$\frac{\alpha_2-\alpha_1}{(p+\alpha_1)(p+\alpha_2)}$	$\frac{\alpha_2-\alpha_1}{(i\omega+\alpha_1)(i\omega+\alpha_2)}$
$\frac{p+\alpha}{(p+\alpha)^2+\omega_0^2}$	$\frac{i\omega+\alpha}{(\omega_0^2-\omega^2)+i2\alpha\omega+\alpha^2}$

Graph of oscillation $s(t)$	Definition of function $s(t)$
	$s(t)=\begin{cases} e^{-\alpha t}\sin\omega_0 t & \text{when } t\geqslant 0, \\ 0 & \text{when } t<0 \end{cases}$ $\alpha>0$
	$s(t)=\begin{cases} \cos\omega_0 t & \text{when } t\geqslant 0 \\ 0 & \text{when } t<0 \end{cases}$
	$s(t)=\begin{cases} \sin\omega_0 t & \text{when } t\geqslant 0 \\ 0 & \text{when } t<0 \end{cases}$

2.14. EXPANSION OF BAND-LIMITED SIGNALS IN FUNCTIONS OF THE FORM SINC (x)

The Kotelnikov sampling theorem, which is widely used in the theory and practice of signals, reads: if the highest frequency in the spectrum of a function $s(t)$ is lower than f_m, the function $s(t)$ is uniquely determined by a sequence of its magnitudes (samples) at the instants of time spaced not more than $1/2\ f_m$ seconds apart.

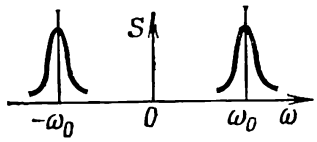
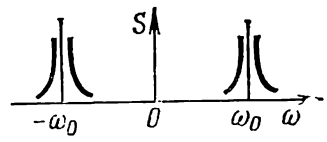
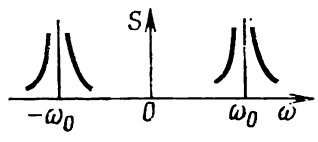
In accordance with this theorem, a signal $s(t)$ whose spectrum is limited by the highest frequency $\omega_m = 2\pi f_m$ may be represented as a series

$$s(t)=\sum_{n=-\infty}^{\infty} s\left(\frac{n}{2f_m}\right)\frac{\sin\omega_m(t-n/2f_m)}{\omega_m(t-n/2f_m)}=\sum_{n=-\infty}^{\infty} s(n\Delta t)\,\varphi_n(t) \quad (2.120)$$

In this expression $1/2f_m = \Delta t$ is the interval between two sampling points on the time axis, and $s(n/2f_m) = s(n\Delta t)$ are the samples of the function $s(t)$ at the instants $t = n\Delta t$.

The representation of a given function $s(t)$ by the series (2.120) is illustrated in Fig. 2.35.

Continued

Laplace transform	Spectral density S (iω)	Graph of modulus S (ω)
$\frac{\omega_0}{(p+\alpha)+\omega_0^2}$	$\frac{\omega_0}{(\omega_0^2-\omega^2)+i2\alpha\omega+\alpha^2}$	
$\frac{p}{p^2+\omega_0^2}$	$\frac{i\omega}{\omega_0^2-\omega^2}$	
$\frac{\omega_0}{p^2+\omega_0^2}$	$\frac{\omega_0}{\omega_0^2-\omega^2}$	

The function of the form

$$\varphi_n(t) = \frac{\sin \omega_m (t - n\Delta t)}{\omega_m (t - n\Delta t)} \tag{2.121}$$

that we have come across earlier (see Sec. 2.9, Fig. 2.20*a*) is called *sampling function* and possesses the following properties:

(*a*) at the point $t = n\Delta t$ $\varphi_n(n\Delta t) = 1$, while at the points $t = k\Delta t$, where k is any positive or negative integer other than n, $\varphi_n(k\Delta t) = 0$;

(*b*) the spectral density of the function $\varphi_0(t)$ is uniform in the frequency band $|\omega| < \omega_m$ and is equal to $1/2f_m = \pi/\omega_m$ [see (2.82) and Fig. 2.20*b*)]. Since the function $\varphi_n(t)$ differs from $\varphi_0(t)$ only in that it is shifted along the time axis by $n\Delta t$, the spectral density of the function $\varphi_n(t)$ is

$$\Phi_n(\omega) = \begin{cases} \frac{1}{2f_m} e^{-in\Delta t\omega} = \Delta t e^{-in\Delta t\omega} & \text{for} -\omega_m < \omega < \omega_m \\ 0 \text{ for } \omega < -\omega_m \text{ and } \omega > \omega_m \end{cases} \tag{2.122}$$

The modulus of this function is shown in the lower part of Fig. 2.36 (solid line).

The fact that the series (2.120) uniquely determines the given signal $s(t)$ at the sampling points does not require additional arguments to prove it, because the samples of the function, i.e., the values $s(n\Delta t)$, themselves are the coefficients of this series. We may show that the series (2.120) determines the function $s(t)$ at any instant t and not only at the sampling points $t = n\Delta t$. For this

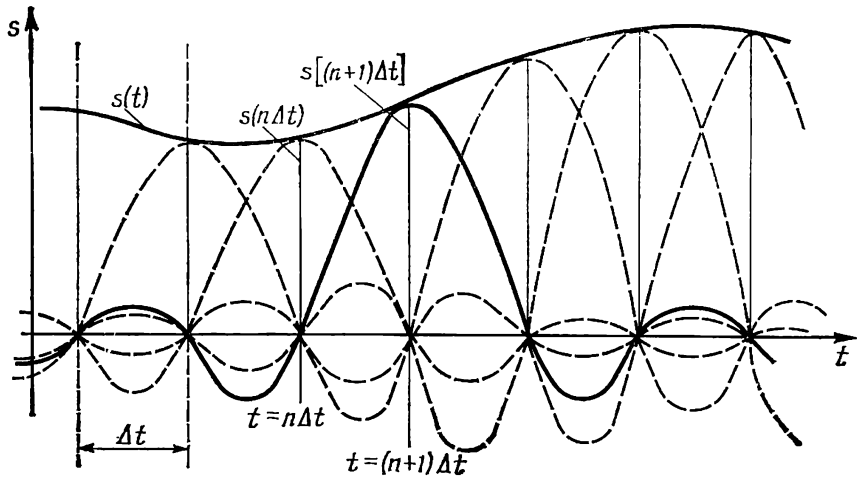


Fig. 2.35. Representation of a signal by the Kotelnikov series

purpose, let us use the general rules of orthogonal-function expansion given in Section 2.2. In this case the expansion is made in terms of the functions of the form (2.121), for which the interval of orthogonality is infinity, while the norm $\| \varphi_n \|$, in accordance with (2.5), is

$$\| \varphi_n \|^2 = \int_{-\infty}^{\infty} \frac{\sin^2 \omega_m (t - n\Delta t)}{\omega_m^2 (t - n\Delta t)^2} \, dt = \frac{1}{\omega_m} \int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2} \, dx = \frac{\pi}{\omega_m} = \Delta t \qquad (2.123)$$

Without predetermining the coefficients of the series (2.120), let us find them by using the general formula (2.9) which holds for the generalized Fourier series:

$$c_n = \frac{1}{\Delta t} \int_{-\infty}^{\infty} s(t) \varphi_n(t) \, dt \qquad (2.124)$$

In so doing, we proceed from the condition that $s(t)$ is a quadratically integrable function (the signal energy is finite).

To calculate the integral in expression (2.124), let us use formula (2.63) according to which

$$\int_{-\infty}^{\infty} s(t) \varphi_n(t) \, dt = \frac{1}{2f_m} \frac{1}{2\pi} \int_{-\omega_m}^{\omega_m} \mathbf{S}(\omega) e^{in\Delta t \omega} \, d\omega \qquad (2.125)$$

In this case the integration limits are brought into correspondence with the given limiting frequency $\omega_m = 2\pi f_m$ in the spectrum of the signal and also in the spectrum of the function $\varphi_n(t)$.

The integral on the right-hand side of (2.125) taken with the factor $1/2\pi$ is none other than the value of $s(t)$ at the moment $t = n\Delta t$. Thus, we have

$$\int_{-\infty}^{\infty} s(t)\,\varphi_n(t)\,dt = \Delta t s(n\Delta t)$$

Substituting this result into (2.124), we obtain the final expression

$$c_n = s(n\Delta t)$$

from which it follows that the coefficients of the series (2.120) are the samples of the function $s(t)$ at the points $t = n\Delta t$.

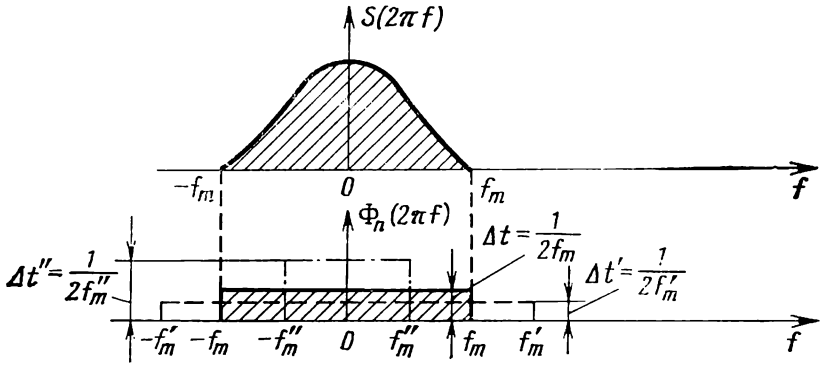


Fig. 2.36. Relation between the spectrum of a signal $s(t)$ and the spectrum of the basic function $\varphi_n(t)$

Since the limitation of the spectrum by a finite highest frequency provides for the continuity of the function $s(t)$, the series (2.120) converges to the function $s(t)$ at any value of t.

If we take a shorter sampling interval $\Delta t'$ ($\Delta t' < \Delta t = 1/2f_m$), the width $2f'_m$ of the spectrum $\Phi'_n(\omega)$ of the function φ'_n will be greater than the width of the spectrum $S(\omega)$ of the signal $s(t)$ (see Fig. 2.36), but this will not affect the coefficients c_n. The modulus of the function $\Phi'_n(\omega)$ is shown in Fig. 2.36 by a dotted line.

Now, if we take $\Delta t''$ longer than Δt, the spectrum $\Phi''_n(\omega)$ of the function $\varphi''_n(t)$ (shown by a dash-and-dot line in Fig. 2.36) will become narrower than the spectrum of the signal $s(t)$, and when calculating the integral in expression (2.125) the integration limits must be taken between $-2\pi f''_m$ and $2\pi f''_m$ instead of between $-2\pi f_m$ and $2\pi f_m$. In this case the coefficients c_n are the samples of some other

function $s_1(t)$, whose spectrum is limited by the highest frequency f'_m, and not of the given signal $s(t)$.

Thus, reducing the sampling interval below $1/2f_m$ is admissible but useless. On the other hand, increasing the interval in excess of $1/2f_m$ is inadmissible.

Now, let us consider a case where the duration of the signal $s(t)$ is finite and equal to T_s, while the frequency band is again equal to f_m. Strictly speaking, these conditions are incompatible, because theoretically, a function of finite duration has an infinitely wide spectrum. In practice, however, we can always define the highest frequency f_m of the spectrum so that the "tails" of the signal, resulting from the cutoff of the frequencies exceeding f_m, contain a negligible fraction of energy as compared with the energy of the original signal $s(t)$. Under such an assumption, if we have a signal of duration T_s and with a frequency band f_m, the total number of independent parameters [i.e., samples $s(n\Delta t)$], which is required for complete specification of the signal, will evidently be

$$N = T_s/\Delta t = 2f_m T_s \tag{2.126}$$

In this case the expression (2.120) assumes the following form (when the time origin is taken to be the first sampling point):

$$s(t) = \sum_{n=0}^{2f_m T_s} s(n\Delta t) \frac{\sin \omega_m (t - n\Delta t)}{\omega_m (t - n\Delta t)} \tag{2.127}$$

The number N is sometimes called the *number of degrees of freedom* of the signal $s(t)$, since even with arbitrarily selected samples $s(n\Delta t)$, the sum (2.127) determines a function fitting the given spectrum and duration of the signal. The number N is also called the *base* of the signal.

The energy and average power of the signal can easily be expressed through a given sequence of time samples.

Using the formulas (2.16) and (2.123), we obtain

$$E = \sum_{n=0}^{2f_m T_s} [s(n\Delta t)]^2 \| \varphi_n \|^2 = \Delta t \sum_{n=0}^{2f_m T_s} [s(n\Delta t)]^2 \tag{2.128}$$

$$\overline{s^2(t)} = \frac{E}{T_s} = \frac{\Delta t}{T_s} \sum_{n=0}^{2f_m T_s} [s(n\Delta t)]^2 = \frac{1}{2f_m T_s} \sum_{n=0}^{2f_m T_s} [s(n\Delta t)]^2 \tag{2.129}$$

From the last expression it is clear that the average (over the time T_s) power of a continuous signal is equal to the mean square of $2f_m T_s$ samples.

2.15. SAMPLING THEOREM IN THE FREQUENCY DOMAIN

Sometimes it is necessary to represent a signal in terms of *frequency samples* of the spectral function $S(\omega)$ rather than in those of time samples of the function $s(t)$. For the function $S(\omega)$ we may

set up a series similar to expression (2.120). This can easily be done on the basis of the interchangeability of the variables t and ω in the Fourier transforms. As applied to expression (2.120), this means that t should be replaced by ω, $2\omega_m$ by T_s, $2f_m$ by $T_s/2\pi$, and $\Delta t = 1/2f_m$ by $\Delta\omega = 2\pi/T_s$.

Thus, we obtain

$$S(\omega) = \sum_{n=-f_m T_s}^{f_m T_s} S(n\Delta\omega) \frac{\sin \frac{T_s}{2}(\omega - n\Delta\omega)}{\frac{T_s}{2}(\omega - n\Delta\omega)}$$

$$= \sum_{n=-f_m T_s}^{f_m T} S\left(n\frac{2\pi}{T_s}\right) \frac{\sin \frac{T_s}{2}\left(\omega - n\frac{2\pi}{T_s}\right)}{\frac{T_s}{2}\left(\omega - n\frac{2\pi}{T_s}\right)} \qquad (2.130)$$

The arrangement of the frequency samples is illustrated in Fig. 2.37.

If in the preceding case the time interval between two adjacent samples had not to exceed $2\pi/2\omega_m$, in the present case the frequency

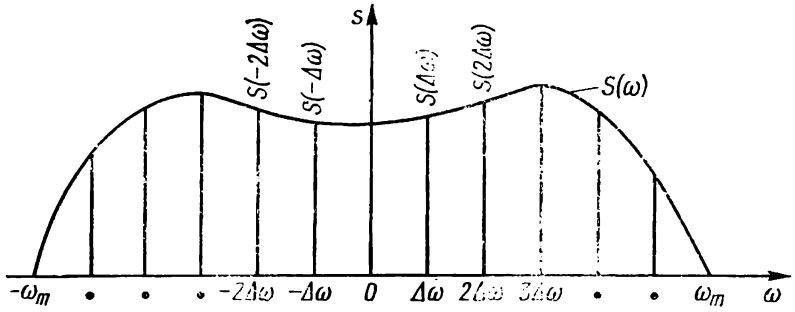


Fig. 2.37. Sampling a signal spectrum after Kotelnikov

interval should not exceed $2\pi/T_s$. With a frequency spectrum $2\omega_m$ wide, embracing the frequency range $-\omega_m < \omega < \omega_m$, the number of samples is equal to $2\omega_m/\Delta\omega = 2f_m T_s$, i.e., it is the same as in the case of a signal represented by the series (2.127).

In the general case, the samples $S(n2\pi/T_s)$ are complex numbers and at each sampling point on the frequency axis there must be specified two parameters—the real and imaginary parts (or the modulus and argument) of $S(n2\pi/T_s)$. Thus, the total number of parameters is twice as large as in the case of the time representation of the signal, where the samples $s(n/2f_m)$ are real numbers. This superfluity of the signal representation in the frequency domain is easily eliminated, because $S(n2\pi/T_s)$ and $S(-n2\pi/T_s)$ are complex conjugates and one of them unambiguously determines the other. Thus, the signal spectrum is completely characterized by a combination of *complex* samples taken solely in the positive frequency

region, and the number of *independent parameters* or *degrees of freedom of the signal* is equal to $2f_mT_s$, as in the case of the signal representation in the time domain.

2.16. CORRELATION ANALYSIS OF DETERMINISTIC SIGNALS

Along with the analysis of signal spectra, it is often necessary in practice to have a characteristic that would give an idea of some properties of the signal, particularly of its rate of change and duration, without the need to resolve it into harmonic components.

Widely used as such a *time characteristic* is the *autocorrelation function* of the signal.

For a deterministic signal $s(t)$ of finite length the autocorrelation function is defined by the following expression

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t)\, s^*(t+\tau)\, dt \qquad (2.131)$$

where τ is the time shift of the signal.

This chapter deals with signals which are real functions of time, so the symbol of the complex conjugate may be deleted:

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t)\, s(t+\tau)\, dt \qquad (2.132)$$

From expression (2.132) it is clear that $B_s(\tau)$ characterizes the degree of connection (correlation) between a signal $s(t)$ and its replica shifted along the time axis by an amount τ. It is also clear that the function $B_s(\tau)$ has a maximum at $\tau = 0$, because any signal is completely correlated with itself. In this case

$$B_s(0) = \int_{-\infty}^{\infty} s^2(t)\, dt = E \qquad (2.133)$$

i.e., the maximum value of the autocorrelation function is equal to the energy of the signal.

As τ increases, the function $B_s(\tau)$ decreases (not necessarily monotonically) and vanishes when the relative shift of the signals $s(t)$ and $s(t+\tau)$ exceeds the signal duration.

Figure 2.38 illustrates the construction of the graph of the autocorrelation function of a simple signal — a square pulse shown in Fig. 2.38*a*. The signal $s(t+\tau)$ advanced by τ is shown in Fig. 2.38*b*, while the product $s(t)\, s(t+\tau)$ is shown in Fig. 2.38*c*. The graph of the function $B_s(\tau)$ is shown in Fig. 2.38*d*. For each τ there are a corresponding product $s(t)\, s(t+\tau)$ and an area under the graph of the function $s(t)\, s(t+\tau)$. The numerical values of such areas for the corresponding τ yield the ordinates of the function $B_s(\tau)$.

Figure 2.39 shows a similar graph for a triangular pulse. From the general definition of the autocorrelation function and the above examples it is evident that it does not matter in which direction —

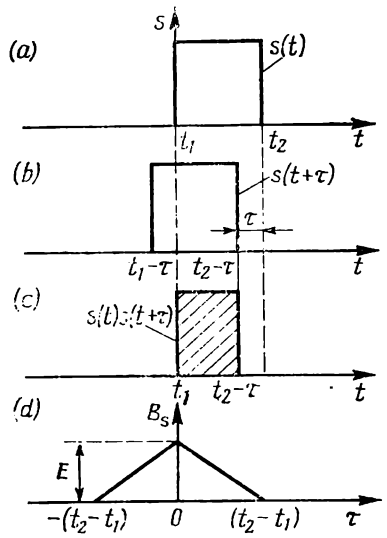


Fig. 2.38. Plotting the autocorrelation function for a square pulse

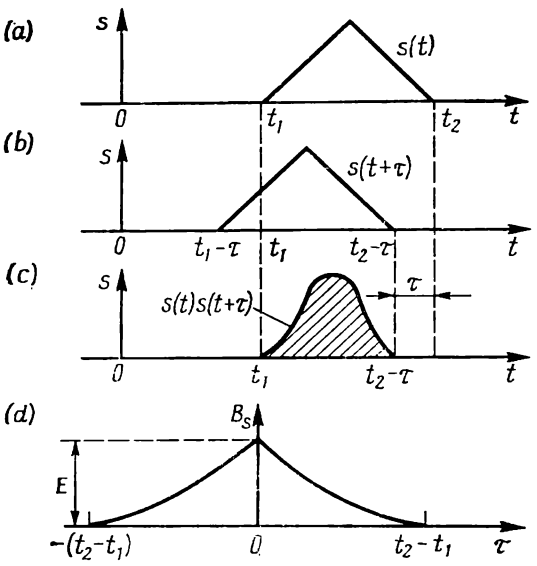


Fig. 2.39. Plotting the autocorrelation function for a triangular pulse

to the right or left—the signal is shifted with respect to its replica. Therefore, expression (2.132) may be generalized in the following way:

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t)\, s(t+\tau)\, dt = \int_{-\infty}^{\infty} s(t)\, s(t-\tau)\, dt \qquad (2.132')$$

This is equivalent to the statement that $B_s(\tau)$ is an *even function of* τ.

Figure 2.40*a* shows a signal in the form of a train of four identical pulses shifted by a time T_1 with respect to one another, and Fig. 2.40*b* presents the autocorrelation function of this signal. In the vicinity of τ equal to 0, $\pm T_1$, $\pm 2T_1$, $\pm 3T_1$ this function is similar to the autocorrelation function of a single pulse (see Fig. 2.39*d*). The maximum value (at $\tau = 0$) of the autocorrelation function of the train is equal to the quadruple energy of a single pulse.

For a periodic signal whose energy is infinitely large the definition of the autocorrelation function by expressions (2.132) and

(2.132′) is unacceptable. Instead, the following definition is used:

$$B_{s,per}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}s(t)\,s(t+\tau)\,dt$$

$$=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}s(t-\tau)\,s(t)\,dt \qquad (2.134)$$

In this case the autocorrelation function assumes the dimension of power, $B_{s,\mathrm{per}}(0)$ being equal to the average power of the periodic

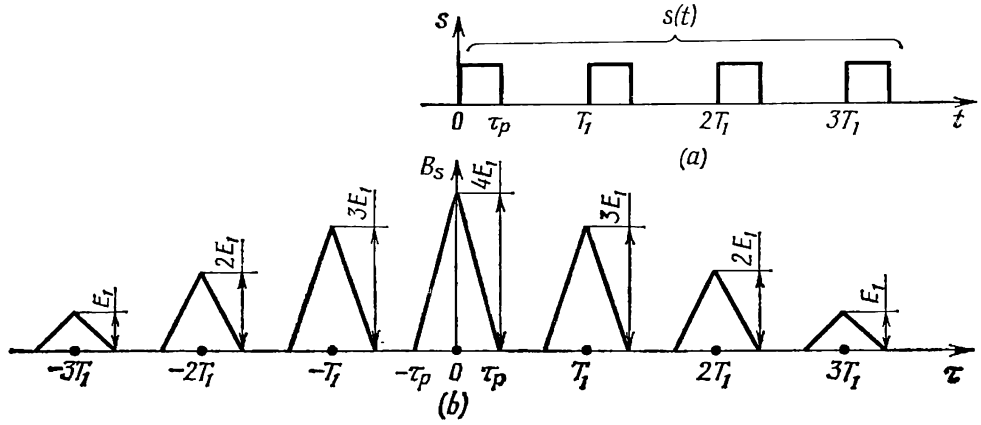


Fig. 2.40. (*a*) Train of four square pulses and (*b*) its autocorrelation function

signal. Because of the periodicity of the signal $s(t)$, the averaging of the product $s(t)\,s(t+\tau)$ or $s(t)\,s(t-\tau)$ over an infinitely large interval T must be in agreement with the averaging over the period T_1. Hence, expression (2.134) may be replaced by the following one:

$$B_{s,\,per}(\tau)=\frac{1}{T_1}\int_{-T_1/2}^{T_1/2}s(t)\,s(t+\tau)\,dt=\frac{1}{T_1}\int_{-T_1/2}^{T_1/2}s(t)\,s(t-\tau)\,dt \qquad (2.135)$$

The integrals in this expression are none other than the autocorrelation function of the signal within the interval T_1. Denoting it by $B_{s\,T_1}(\tau)$, we come to the relation

$$B_{s,\,per}(\tau)=B_{s\,T_1}(\tau)/T_1 \qquad (2.136)$$

From expression (2.135) follows another obvious statement: the autocorrelation function, $B_{s,per}(\tau)$ of a periodic signal $s(t)$ is also periodic. The period of the function $B_{s,per}(\tau)$ coincides with the period T_1 of the original signal $s(t)$. For example, the autocorrelation function of a simple (harmonic) oscillation $s(t)=$

$= A_0 \cos(\omega_0 t + \theta_0)$ is

$$B_{s,\,per}(\tau) = \frac{A_0^2}{T_1} \int_{-T_1/2}^{T_1/2} \cos(\omega_0 t + \theta_0) \cos[\omega_0 (t+\tau) + \theta_0]\, dt$$

$$= \frac{1}{2} A_0^2 \cos \omega_0 \tau, \quad \omega_0 = \frac{2\pi}{T_1} \tag{2.137}$$

At $\tau = 0$, $B_{s,per}(0) = 1/2\, A_0^2$ is the average power of a harmonic oscillation with an amplitude A_0. It should be emphasized that the

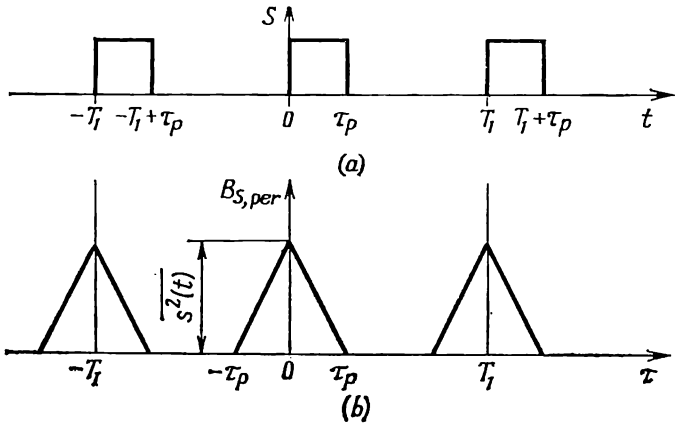


Fig. 2.41. (*a*) Periodic train of pulses and (*b*) its autocorrelation function

autocorrelation function $B_{s,per}(\tau)$ is independent of the epoch angle θ_0 of the oscillation.

Figure 2.41 shows the autocorrelation function of a periodic train of square pulses (Fig. 2.41*a*). Each of the pulses of the function $B_{s,per}(\tau)$ has the same shape as the autocorrelation function of a single pulse in the periodic train $s(t)$. However, in this case the maximum ordinates of $B_{s,per}(\tau)$ are equal to the average power $\overline{s^2(t)}$ of the signal $s(t)$ and not to its energy (as in Fig. 2.40).

The degree of connection between two different signals $s_1(t)$ and $s_2(t)$ is estimated by means of the *cross-correlation function* defined by the general expression

$$B_{s_1 s_2}(\tau) = \int_{-\infty}^{\infty} s_1(t)\, s_2^*(t+\tau)\, dt \tag{2.138}$$

For real functions $s_1(t)$ and $s_2(t)$

$$B_{s_1 s_2}(\tau) = \int_{-\infty}^{\infty} s_1(t)\, s_2(t+\tau)\, dt \tag{2.139}$$

The above discussed autocorrelation function $B_s(\tau)$ is a particular case of the function $B_{s_1 s_2}(\tau)$, where $s_1(t) = s_2(t)$.

An example of the cross-correlation function for two signals $s_1(t)$ and $s_2(t)$ is given in Fig. 2.42. The initial position of the signals ($\tau = 0$) is shown in Fig. 2.42*a*. As the signal $s_2(t)$ is shifted to the left ($\tau > 0$, Fig. 2.42*b*), the cross-correlation function first increases and then decreases and vanishes at $\tau = T$. As the signal is shifted to the right ($\tau < 0$), the cross-correlation function immediately decreases. As a result, the function $B_{s_1 s_2}(\tau)$ (Fig. 2.42*c*) is

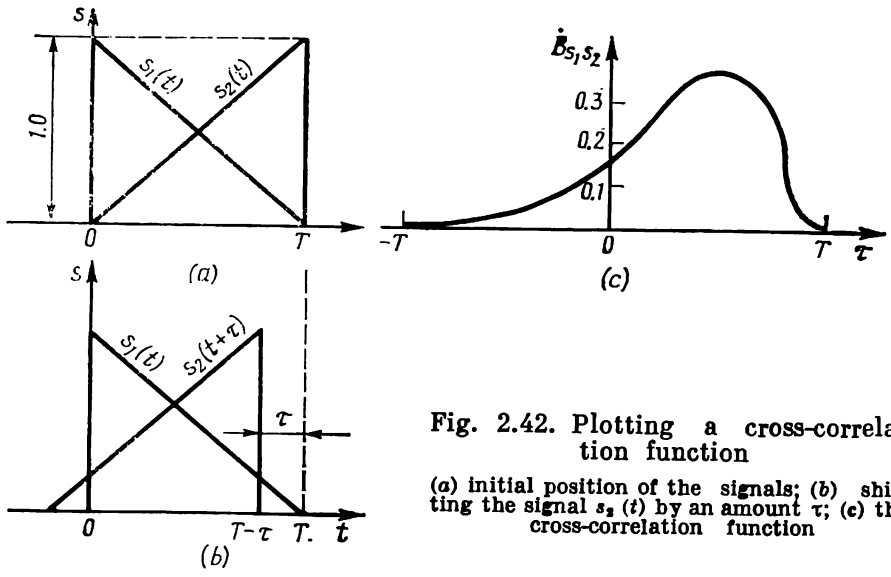


Fig. 2.42. Plotting a cross-correlation function

(*a*) initial position of the signals; (*b*) shifting the signal $s_2(t)$ by an amount τ; (*c*) the cross-correlation function

asymmetrical relative to the axis of ordinates.

It is clear that the value of $B_{s_1 s_2}$ will not change if one delays the signal $s_1(t)$ rather than advances the signal $s_2(t)$. Therefore, expression (2.139) can be generalized in the following way:

$$B_{s_1 s_2}(\tau) = \int_{-\infty}^{\infty} s_1(t)\, s_2(t+\tau)\, dt = \int_{-\infty}^{\infty} s_1(t-\tau)\, s_2(t)\, dt \quad (2.140)$$

A distinction should be made, however, between expressions (2.132′) and (2.140). In contrast to $B_s(\tau)$, the cross-correlation function is *not necessarily* even with respect to τ. Furthermore, the cross-correlation function is *not necessarily* maximum at $\tau = 0$. Both these properties of the function $B_{s_1 s_2}(\tau)$ are illustrated in Fig. 2.42.

2.17. RELATIONSHIP BETWEEN THE AUTOCORRELATION FUNCTION AND THE SPECTRAL CHARACTERISTIC OF A SIGNAL

Let us use expression (2.63), where we set $f(t) = s(t)$, $g(t) = s(t+\tau)$ and, respectively, $\mathbf{F}(\omega) = \mathbf{S}(\omega)$ and $\mathbf{G}(\omega) = \mathbf{S}(\omega)\, e^{i\omega\tau}$. In this case we obtain

$$\int_{-\infty}^{\infty} s(t)\, s(t+\tau)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, \mathbf{S}^*(\omega)\, e^{-i\omega\tau}\, d\omega = B_s(\tau)$$

Taking into account that $\mathbf{S}(\omega)\, \mathbf{S}^*(\omega) = S^2(\omega)$, we obtain the relationship sought:

$$B_s(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S^2(\omega)\, e^{-i\omega\tau}\, d\omega \qquad (2.141)$$

On the basis of the known properties of the Fourier transforms, we may also write*

$$S^2(\omega) = \int_{-\infty}^{\infty} B_s(\tau)\, e^{i\omega\tau}\, d\tau \qquad (2.142)$$

Thus, the direct Fourier transform (2.142) of the autocorrelation function $B_s(\tau)$ yields the spectral energy density (see the note at the end of Sec. 2.8), while the transform (2.141) yields the autocorrelation function $B_s(\tau)$.

From expressions (2.141) and (2.142) one may infer properties similar to those mentioned in Sec. 2.10: the wider the spectrum $S(\omega)$ of a signal, the shorter the correlation interval, i.e., the shift τ, within which the autocorrelation function of the signal is nonzero. Correspondingly, the longer the correlation interval of the given signal, the narrower its spectrum.

From expressions (2.141) and (2.142) it is also clear that the autocorrelation function $B_s(\tau)$ is independent of the phase characteristic of the signal spectrum. Since the form of the function $s(t)$ of a given amplitude spectrum $S(\omega)$ depends essentially on the phase spectrum, we may draw the following conclusion: signals $s(t)$ of different forms but of the same amplitude spectrum will have the same autocorrelation function $B_s(\tau)$.

2.18. COHERENCE

In physics, where it was used for the first time, the term "coherence" implied the coincidence of the phases of harmonic oscillations subject to summing. At present, in radio engineering and in-

* Owing to the fact that the function $B_s(\tau)$ is even, the sign before $i\omega\tau$ in the exponent is of no importance. The same applies to (2.141).

formation theory, the term is treated in a wider sense: by coherence is usually meant the relation between the phases of signals.

The degree of coherence of signals may be estimated by comparing the energy of the sum of the signals with the sum of the energies of the individual component signals.

Let us consider two signals, $s_1(t)$ and $s_2(t)$. The energy of their sum in the general case is

$$E = \int_{-\infty}^{\infty} [s_1(t) + s_2(t)]^2 dt = \int_{-\infty}^{\infty} [s_1(t)]^2 dt + \int_{-\infty}^{\infty} [s_2(t)]^2 dt + 2\int_{-\infty}^{\infty} s_1(t)\, s_2(t)\, dt \quad (2.143)$$

The first two integrals on the right-hand side of this equation define the energies E_1 and E_2 of the signals $s_1(t)$ and $s_2(t)$ taken separately, while the last integral defines the "energy of interaction" E_{12} between these signals.

In terms of the symbols used in the correlation analysis of signals (see Secs. 2.16 and 2.17), expression (2.43) assumes the form

$$E = B_s(0) + B_{s_2}(0) + 2B_{s_1 s_2}(0) \quad (2.144)$$

Thus, the cross-correlation function $B_{s_1 s_2}(\tau)$, introduced in Sec. 2.16 [see (2.139)], at $\tau = 0$ may serve as a measure of the interaction energy E_{12}.

The greater the absolute contribution of the interaction energy E_{12} to the total energy E, the stronger the coherence of the signals $s_1(t)$ and $s_2(t)$.

Sometimes, another criterion, namely, the ratio

$$K = 2\int_{-\infty}^{\infty} s_1(t)\, s_2(t)\, dt \Big/ 2\int_{-\infty}^{\infty} |s_1(t)|\, |s_2(t)|\, dt \quad (2.145)$$

is used as a measure of coherence.

In this expression, the numerator is the interaction energy E_{12}, and the magnitude of the ratio can be anything between -1 and $+1$.

As an example, let us consider two identical signals whose energies E_1 and E_2 are equal. When these signals are added in phase, the total energy $E = 4E_1 = 4E_2$, while $E_{12} = 2E_1 = 2E_2$. In this case the denominator of the fraction on the right-hand side of (2.145) is equal to $E_1 + E_2$ (like the numerator), and $K = +1$. If the signals are added in antiphase (i.e., in the case of their subtraction), $E = 0$, and $K = -1$.

In both cases the signals are completely coherent. In order for the signals to be incoherent (i.e., for K to be equal to zero), the following condition must be satisfied:

$$\int_{-\infty}^{\infty} s_1(t)\, s_2(t)\, dt = 0 \qquad (2.146)$$

But this condition is none other than the condition of orthogonality of the signals in question, so we arrive at a conclusion that incoherent signals are purely orthogonal. From the orthogonality of incoherent signals it follows that when such signals are added linearly, the energy of their sum is equal to the sum of the energies of the individual component signals.

Chapter 3

RADIO SIGNALS

3.1. GENERAL

To convey information over great distances, use is made of such signals as can be effectively radiated by antenna systems, simultaneously possessing characteristics to enable their passage as free radio waves through the transmission medium between the transmitting end and the receiving point. These are *radio-frequency oscillations.* The information to be conveyed must, by some means or other, be placed on a radio-frequency oscillation known as the *carrier.* The frequency ω_0 of this oscillation is selected depending on the distance over which information is to be transmitted, radio-wave propagation characteristics of the transmission medium, and some other technical and economical factors. But in any case, the frequency ω_0 must be much higher than the uppermost frequency Ω_m in the spectrum of the message to be transmitted*.

This requirement stems from the fact that to ensure undistorted transmission of a message through a radio system (the transmission medium included), the width of the spectrum Ω_m of this message must be narrow as compared with the carrier frequency ω_0, for the smaller the ratio Ω_m/ω_0, the less prominent the effect of imperfections of the system characteristics. Therefore, the higher the required speed of transmission of information, and consequently, the wider the spectrum Ω_m of the message to be transmitted, the higher must be the carrier frequency of the radio signal. As a rule, the inequality $\Omega_m/\omega_0 \ll 1$ is satisfied.

Any radio signal may therefore be treated as a "narrow-band" process even when transmitting "wide-band" messages.

Let us consider the following examples. When transmitting speech or music, the message spectrum is usually confined within the frequency range from $F_{\min} = 30$ to 50 Hz to $F_{\max} = 3000$ to 10,000 Hz. Even in the case of the longest broadcasting-band wave of $\lambda = 2000$ m (carrier frequency $f_0 = 150$ kHz), the ratio $F_{\max}/f_0 = 10^4/(1.5 \times 10^5) \approx 0.06$, and when transmitting similar messages on short waves (carrier frequencies from 15 to 20 MHz), this ratio does not exceed a few hundredths of a percent. In television, the frequency band of the message is fairly wide (5 to 6 MHz), but

* In this chapter the symbol Ω is used to denote the frequency of the modulating function (i.e., the message to be transmitted).

the carrier frequency is also taken to be at least 50 to 60 MHz, so that the ratio F_{max}/f_0 does not exceed 10%.

Generally, an information-bearing radio signal may be represented in the form

$$\begin{aligned} a(t) &= A(t) \cos [\omega_0 t + \theta(t)] \\ &= A(t) \cos \psi(t) \end{aligned} \tag{3.1}$$

where the amplitude A or the phase angle θ is varied in accordance with the information to be transmitted.

If A and θ are constants, expression (3.1) describes a simple harmonic oscillation carrying no information. If A or θ (and consequently ψ) is made to vary in order to transmit a message, the oscillation becomes *modulated.*

According to the parameter being changed (amplitude A or phase angle θ), two basic types of modulation are known, namely, *amplitude modulation* and *angle modulation.* The latter is in turn subdivided into two types—*frequency modulation* and *phase modulation.* These two types of modulation are closely related and differ only in the character of variation in time of the angle ψ, the modulating function being the same.

A modulated wave has a spectrum whose structure depends both on the spectrum of the transmitted signal and the type of modulation. The fact that the width of the spectrum of the modulating signal is narrow in comparison with the carrier frequency ω_0 allows one to regard $A(t)$ and $\theta(t)$ as *slowly varying* functions of time. This means that the relative variation of $A(t)$ or $\theta(t)$ during a period of the carrier wave is small compared with unity.

First, let us consider amplitude variation. With the amplitude varying at rate dA/dt, the amplitude increment during period T_0 can be considered to be approximately equal to $(dA/dt)\,T_0$. Hence, the relative change of the amplitude during a period is

$$\left|\frac{dA}{dt}\right| \frac{T_0}{A} = \left|\frac{dA}{dt}\right| \frac{1}{A} \frac{2\pi}{\omega_0}$$

We may presume that the condition of slow variation of the function $A(t)$ will be satisfied if

$$\frac{2\pi}{\omega_0} \left|\frac{dA}{dt}\right| \frac{1}{A} \ll 1 \quad \text{or} \quad \left|\frac{dA}{dt}\right| \frac{1}{A} \ll \frac{\omega_0}{2\pi} \tag{3.2}$$

In a similar way we may define the condition of slow variation of the function $\theta(t)$.

Since the instantaneous frequency of an oscillation is equal to the rate of change of its phase angle (this question is discussed in detail in the following sections of the book), the differentiation of the argument in expression (3.1) yields

$$\omega(t) = \frac{d\psi(t)}{dt} = \omega_0 + \frac{d\theta}{dt}$$

The derivative $d\theta/dt$ defines the deviation of the frequency $\omega(t)$ from the carrier frequency ω_0. This deviation can be either fast or slow. So that the oscillation $a(t)$ may be regarded as being close to a harmonic one, we must require that the change in frequency during a cycle $T_0 = 2\pi/\omega_0$ should be small as compared with the frequency $\omega(t)$ at a given instant of time.

Thus, the condition of slow variation of the function $\theta(t)$ may be written down in the form of the following inequality:

$$\frac{\left|\frac{d}{dt}\left(\frac{d\theta}{dt}\right)\right| T}{\omega(t)} \ll 1 \quad \text{or} \quad \left|\frac{d^2\theta}{dt^2}\right| \ll \frac{\omega(t)}{T}$$

Since the frequency $\omega(t)$ usually differs but little from ω_0, we may write

$$\left|\frac{d^2\theta}{dt^2}\right| \ll \frac{1}{2\pi}\omega_0^2 \tag{3.3}$$

For the majority of radio signals, the inequalities (3.2) and (3.3) are usually satisfied. This means that in any type of modulation, the respective parameter (amplitude, phase or frequency) of a radio signal varies so slowly that within a single period T_0 the wave can be considered to be sinusoidal.

This premise underlies the entire subsequent discussion of the properties of radio signals and their spectra.

3.2. AMPLITUDE-MODULATED RADIO SIGNALS

Amplitude modulation is the simplest and most universally employed method of imparting intelligence to a radio-frequency oscillation (the carrier). In this type of modulation, the envelope of the carrier varies in accordance with the transmitted message, while the frequency and epoch angle of the wave remain unchanged. For an amplitude-modulated radio signal, therefore, the general expression (3.1) can be replaced by the following one:

$$a(t) = A(t)\cos(\omega_0 t + \theta) \tag{3.4}$$

The character of the envelope $A(t)$ depends on the type of the transmitted signal. When the message is continuous (Fig. 3.1*a*), the modulated wave assumes the form shown in Fig. 3.1*b*. The envelope $A(t)$ is identical with the modulating function, i.e., the message $s(t)$. Figure 3.1*b* is constructed on the premise that the d-c component of the function $s(t)$ is zero (otherwise the carrier amplitude A_0 in modulation may not coincide with the amplitude of the unmodulated wave). The largest "down" change in $A(t)$ cannot be larger than the amplitude A_0 of the carrier, while the "up" change, in principle, can exceed A_0.

The basic parameter of an amplitude-modulated wave is the *degree of modulation* stated in terms of the *modulation factor* (usually expressed in percent, in which case it is called *percentage modulation*).

The definition of this motion is particularly graphic in the case of tone modulation, where the modulating function $s(t)$ is a harmonic oscillation:

$$s(t) = S_0 \cos(\Omega t + \gamma)$$

In this case the envelope of the modulated wave may be written as

$$A(t) = A_0 + k_{am}s(t) = A_0 + \Delta A_m \cos(\Omega t + \gamma) \qquad (3.5)$$

where Ω is the modulation frequency; γ, the epoch angle of the envelope; k_{am}, a proportionality factor; and $\Delta A_m = k_{am}S_0$, the amplitude of the envelope variation (Fig. 3.2).

The ratio

$$M = \Delta A_m / A_0$$

is the *modulation factor*.

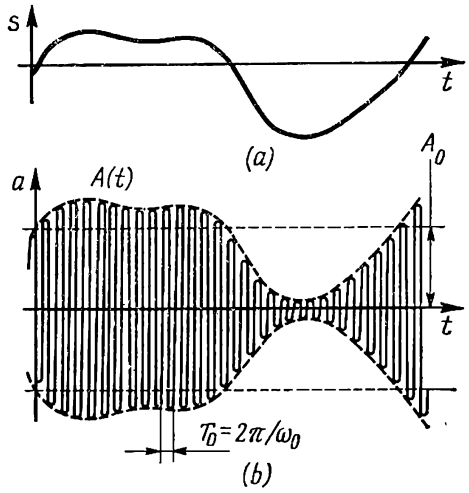


Fig. 3.1. (*a*) Modulating function and (*b*) amplitude-modulated oscillation

Thus, the instantaneous value of the modulated wave can be written in the form

$$a(t) = A_0 [1 + M \cos(\Omega t + \gamma)] \cos(\omega_0 t + \theta_0) \qquad (3.6)$$

In the case of undistorted modulation ($M \leqslant 1$), the amplitude of the oscillation varies from the minimum $A_{\min} = A_0(1 - M)$ to the maximum $A_{\max} = A_0 \times (1 + M)$.

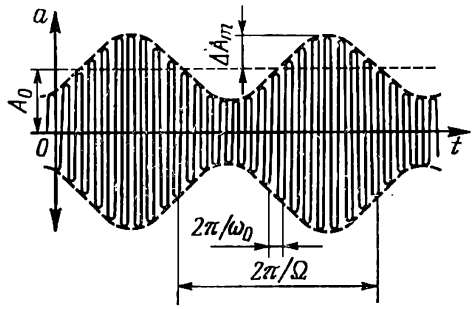


Fig. 3.2. Oscillation modulated in amplitude by a harmonic function

The average power of the modulated wave over a period of the high-frequency carrier varies in accordance with the amplitude variations. The power corresponding to the peaks of the envelope is by a factor of $(1 + M)^2$ higher than the power of the carrier, but the average power over a modulation period is proportional to the mean* square of the amplitude $A(t)$:

$$\overline{A^2(t)} = A_0 \overline{[1 + M \cos(\Omega t + \gamma)]^2} = A_0^2(1 + 0.5M^2) \qquad (3.7)$$

* The mean value of $\cos(\Omega t + \gamma)$ over a period of the modulation frequency is zero, while the mean value of $\cos^2(\Omega t + \gamma)$ is equal to 1/2. The line above the function denotes time average.

This power exceeds the carrier power only by a factor of $(1 + 0.5M^2)$. Thus, in the case of a 100-percent modulation ($M = 1$), the peak power is equal to $4P_0$, while the average power is equal to $1.5P_0$, $P_0 = (1/2)A_0^2$ being the power in the carrier. From this it follows that the increase in the wave power due to modulation (this increase chiefly determining the message detection conditions at the receiving end) does not exceed one half of the power in the carrier wave even in the case of a 100-percent modulation.

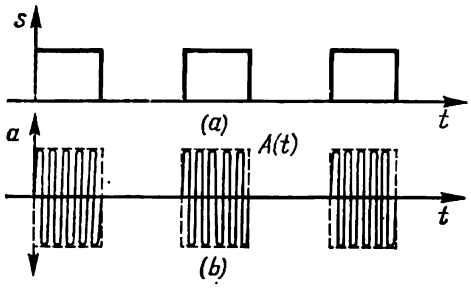


Fig. 3.3. Oscillation modulated in amplitude by a sequence of pulses

When transmitting discrete messages comprising a sequence of pulses (Fig. 3.3*a*), the modulated waves assume the form of a train of radio pulses shown in Fig. 3.3*b*.

In this case it is presumed that the epoch angles of the radio-frequency carrier in each pulse of the modulated train are the same as in the case of gating a single continuous harmonic wave. It is on this premise only that the train of radio pulses shown in Fig. 3.3*b* can be regarded as a wave modulated solely in amplitude. If the phase angle changes from pulse to pulse, one should speak of a combined, amplitude-angle modulation.

3.3. FREQUENCY SPECTRUM OF AN AMPLITUDE-MODULATED SIGNAL

Let a modulated radio-frequency oscillation be given, whose frequency ω_0 and epoch angle θ_0 are constants, and whose envelope $A(t)$ bears a signal (or message) $s(t)$ being transmitted. Such an oscillation can be represented analytically by expression (3.4).

It is required to find the relation between the spectrum of the modulated wave and that of the modulating function, i.e., the spectrum of the original message. A most simple and graphic way to do this is to consider the case of tone (harmonic) modulation where the envelope is given by

$$A(t) = A_0 [1 + M \cos(\Omega t + \gamma)]$$

and the modulated wave is defined by expression (3.6).

Let this expression be rewritten in the form

$$a(t) = A_0 [\cos(\omega_0 t + \theta_0) + M \cos(\Omega t + \gamma) \cos(\omega_0 t + \theta_0)]$$

The second term in the brackets, which is the product of modulation, can be reduced to

$$M\cos(\Omega t+\gamma)\cos(\omega_0 t+\theta_0)=\frac{M}{2}[(\omega_0+\Omega)t+(\theta_0+\gamma)]+\frac{M}{2}\cos[(\omega_0-\Omega)t+(\theta_0-\gamma)]$$

After that, the expanded expression for the wave $a(t)$ assumes the form

$$a(t)+A_0\cos(\omega_0 t+\theta_0)+\frac{MA_0}{2}\cos[(\omega_0+\Omega)t+\theta_0+\gamma]+\frac{MA_0}{2}\cos[(\omega_0-\Omega)t+\theta_0-\gamma] \qquad (3.8)$$

The first term on the right-hand side of this equation represents the original, unmodulated wave of frequency ω_0. The second and third terms correspond to new (harmonic) waves developing in the process of amplitude modulation. The frequencies, $\omega_0+\Omega$ and $\omega_0-\Omega$, of these waves are known as the upper and lower side frequencies.

The amplitudes of these two waves are the same and are equal to $(M/2)A_0$ and their phases are symmetrical with respect to the phase of the carrier wave. This is illustrated by the vector diagram shown in Fig. 3.4. On this diagram, the time axis revolves clockwise with an angular frequency ω_0, the angle $\omega_0 t$ being measured from the line OB. Therefore, the carrier $A_0\cos(\omega_0 t+\theta_0)$ is shown in this diagram as a stationary vector OD having a length A_0 and making an angle θ_0 with the horizontal line OB. The instantaneous value of the carrier oscillation at a time t is equal to the projection OK of the vector A_0 on the time axis (line segment OK).

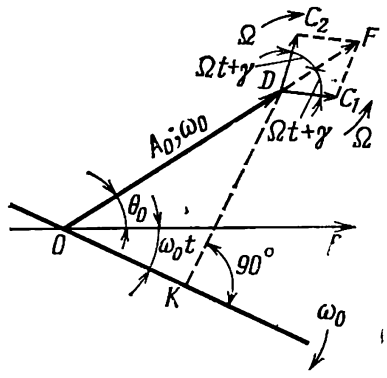


Fig. 3.4. Vectorial representation of an amplitude-modulated oscillation

The oscillation with a frequency $\omega_0+\Omega$ exceeding the angular frequency of revolution of the time axis by an amount Ω can be shown on the same diagram as a vector revolving counterclockwise with an angular frequency Ω (vector DC_1). The oscillation with a frequency $\omega_0-\Omega$ can be represented by a vector revolving clockwise with the same frequency Ω (vector DC_2). Therefore, the oscillations of the upper and lower side frequencies are shown by two vectors having a length $MA_0/2$ and revolving in opposite directions. The phasing of these vectors is symmetrical with respect to the

vector A_0 of the carrier oscillation. This follows from expression (3.8) which, for clarity, can be written as

$$a(t) = A_0 \cos(\omega_0 t + \theta_0) + \frac{MA_0}{2} \cos[(\omega_0 t + \theta_0) + (\Omega t + \gamma)]$$

$$+ \frac{MA_0}{2} \cos[(\omega_0 t + \theta_0) - (\Omega t + \gamma)]$$

From this expression it is evident that with any epoch angle γ of the envelope the vectors DC_1 and DC_2 corresponding to the oscillations of the upper and lower side frequencies are disposed symmetrically about OD and make angles equal to $\pm(\Omega t + \gamma)$ with the vector of the carrier frequency. In Fig. 3.4 the origin of these vectors is transferred from the point O to the point D. The resultant vector DF (which is the vector sum of the vectors DC_1 and DC_2) is called modulation vector and always lies on the line OD, hence, the sum of all the three oscillations, i.e., the carrier and the two side-frequency oscillations, can be regarded as a wave of constant epoch angle and frequency, but of modulated amplitude.

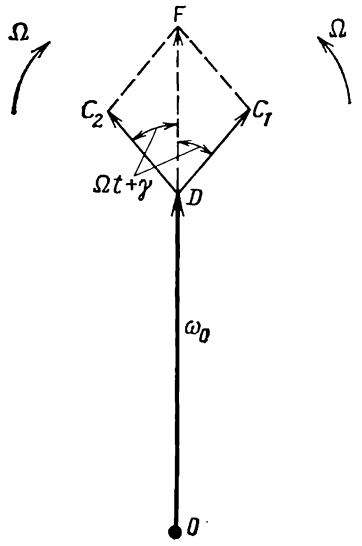


Fig. 3.5. Vector diagram of amplitude modulation with the epoch angle θ_0 of the carrier equal to 90°

In addition, it should be noted that if the equality of the amplitudes of the side-frequency oscillations, or the symmetry of their phases relative to the carrier phase, is disturbed as a result of the passage of the modulated wave through radio circuits, the resultant wave vector will fluctuate with respect to the direction OD. This is equivalent to the development of a parasitic phase modulation.

Now let us consider the phase of the envelope in the case of purely amplitude modulation. Assume that the epoch angle of the carrier is $\theta_0 = 90°$. In this case the vector diagram takes the form shown in Fig. 3.5. If the vectors DC_1 and DC_2 of the side-frequency oscillations are directed upwards when $\Omega t = 0$ (position I in Fig. 3.6), the envelope at this instant will pass through its maximum $A_0(1 + M)$. This case corresponds to the zero epoch angle γ of the envelope [see equation (3.6)], and the equation of the envelope will be

$$A(t) = A_0(1 + M \cos \Omega t)$$

If the vectors DC_1 and DC_2 occupy a horizontal position at the instant $\Omega t = 0$, the envelope passes through A_0. In this case the

epoch angle of the envelope is $\gamma = -\pi/2$, and the equation of the envelope is

$$A(t) = A_0(1 + M \sin \Omega t)$$

The positions of the vectors DC_1 and DC_2 for $\Omega t = \pi/2$, π, and $3\pi/2$ when $\gamma = 0$ are indicated in Fig. 3.6 by the numerals *II*, *III* and *IV*, respectively.

The spectral diagram of a tone-modulated wave is shown in Fig. 3.7. In this case the width of the spectrum is equal to the double

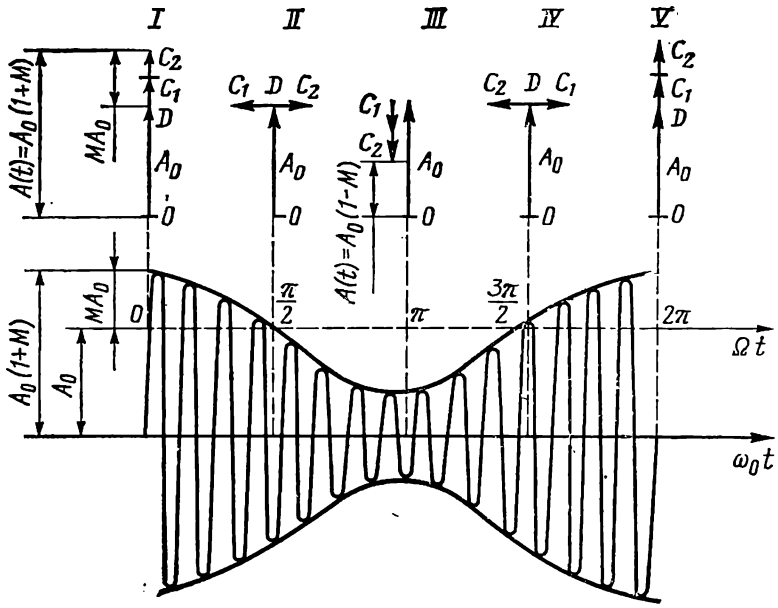


Fig. 3.6. Phasing of the side-frequency oscillations at various instants of time

modulation frequency, 2Ω, and the amplitudes of the side-frequency oscillations cannot exceed half the amplitude of the unmodulated wave (at $M \leqslant 1$).

Similar results can be obtained for modulation by any complex signal. The formation of the spectrum of an amplitude-modulated wave can be explained in a most simple way by taking first an example of a modulating signal $s(t)$ which is a sum of two tones:

$$s(t) = S_1 \cos \Omega_1 t + S_2 \cos \Omega_2 t$$

By analogy with expression (3.5), we may write

$$A(t) = A_0 + \Delta A_{m_1} \cos \Omega_1 t + \Delta A_{m_2} \cos \Omega_2 t$$
$$= A_0 (1 + M_1 \cos \Omega_1 t + M_2 \cos \Omega_2 t)$$

Substituting this expression into equation (3.4) and making some trigonometric transformations similar to those performed for obtaining equation (3.8), we get the following result (for simplicity, the epoch angle of both the carrier wave and the modulating signal is omitted):

$$a(t) = A_0 \cos \omega_0 t + \frac{M_1 A_0}{2} \cos (\omega_0 + \Omega_1) t + \frac{M_1 A_0}{2} \cos (\omega_0 - \Omega_1) t$$

$$+ \frac{M_2 A_0}{2} \cos (\omega_0 + \Omega_2) t + \frac{M_2 A_0}{2} \cos (\omega_0 - \Omega_2) t$$

From this expression it follows that the frequencies Ω_1 and Ω_2 each produce tone modulation of its own, accompanied by the development of a pair of side frequencies, this process being linear in the sense that the amplitudes and phases of the side-frequency oscillations resulting from the different modulating voltages are mutually independent (this property obtains, provided that the total "down" variation of the envelope does not exceed 100%).

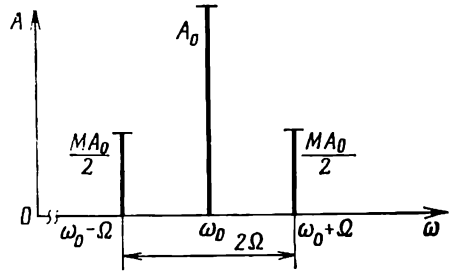


Fig. 3.7. Oscillation spectrum in tone (harmonic) amplitude modulation

From this example one can easily derive a rule for constructing the spectral diagram of an amplitude-modulated oscillation $a(t)$ on the basis of a given spectrum of the modulating function $s(t)$. Let the spectrum of this function have the form shown in Fig. 3.8*a*. In this figure, S_1, S_2, . . ., S_n, . . . stand for the amplitudes of the harmonic oscillations in the spectrum of the message $s(t)$, while Ω_{min} and Ω_{max} denote the limiting frequencies of the spectrum.

The spectral diagram of a radio-frequency oscillation modulated in amplitude by the message $s(t)$ is shown in Fig. 3.8*b*. The modulation factors M_1, M_2, . . ., M_n are proportional to the amplitudes S_1, S_2, . . ., S_n of the corresponding tones entering into the complex message $s(t)$.

Now, consider the general case where the spectrum of the message $s(t)$ is not necessarily discrete. Let us proceed from the general expression (3.4). The message to be transmitted is represented by the variation of the envelope $A(t)$. Without predetermining the form of the function $s(t)$, let us obtain an expression for the spectral density $S_a(\omega)$ of the modulated wave $a(t)$ regarded as the product of the envelope $A(t)$ and the harmonic oscillation $\cos(\omega_0 t + \theta_0)$.

Setting $s(t) = A(t)$ in expression (2.58) we have

$$\mathbf{S}_a(\omega) = \int_{-\infty}^{\infty} A(t) \cos(\omega_0 t + \theta_0) e^{-i\omega t} dt$$
$$= (1/2) e^{i\theta_0} \mathbf{S}_A(\omega - \omega_0) + (1/2) e^{-i\theta_0} \mathbf{S}_A(\omega + \omega_0) \quad (3.9)$$

In this expression $\mathbf{S}_A$ stands for the spectral density of the envelope, i.e., of the modulating function.

It should be noted that the spectrum of a slowly varying function of time, $A(t)$, is concentrated in the region of comparatively low frequencies. Therefore, the function $\mathbf{S}_A(\omega - \omega_0)$ is essentially nonzero only at frequencies ω close to ω_0, i.e., when the difference $\omega - \omega_0 = \Omega$ is relatively small. Similarly, the term $\mathbf{S}_A(\omega + \omega_0)$ exists at frequencies close to $-\omega_0$.

So, the spectral density $\mathbf{S}_a(\omega)$ of the modulated wave forms two spikes: one near $\omega = \omega_0$ and the other near $\omega = -\omega_0$. Therefore, for a narrow-band signal, we may presume that in the region of positive frequencies

$$\mathbf{S}_a(\omega) \approx \frac{1}{2} e^{i\theta_0} \mathbf{S}_A(\omega - \omega_0) \quad (3.10)$$

while in the region of negative frequencies

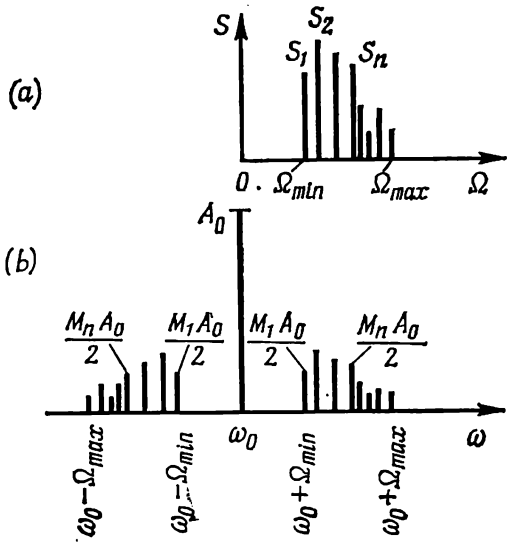


Fig. 3.8. Discrete spectra of (*a*) complex modulating function and (*b*) amplitude-modulated oscillation

$$\mathbf{S}_a(\omega) \approx \frac{1}{2} e^{-i\theta_0} \mathbf{S}_A(\omega + \omega_0) \quad (3.10')$$

To explain the rule for constructing the spectrum $\mathbf{S}_a(\omega)$, let us take the following example. Let the envelope of a radio-frequency oscillation have the form

$$A(t) = A_0 [1 + k_{am} s(t)] \quad (3.11)$$

where $s(t)$ is the message to be transmitted and has a spectral density $\mathbf{S}(\Omega)$, and the factor k_{am} has the same meaning as in expression (3.5).

The spectral density of the envelope $A(t)$ is shown in Fig. 3.9*a*. The discrete component of this spectrum, equal to $2\pi A_0 \delta(\Omega)$ corresponds to the constant A_0, and the continuous component $k_{am} A_0 \mathbf{S}(\Omega)$ corresponds to the message $s(t)$ being transmitted.

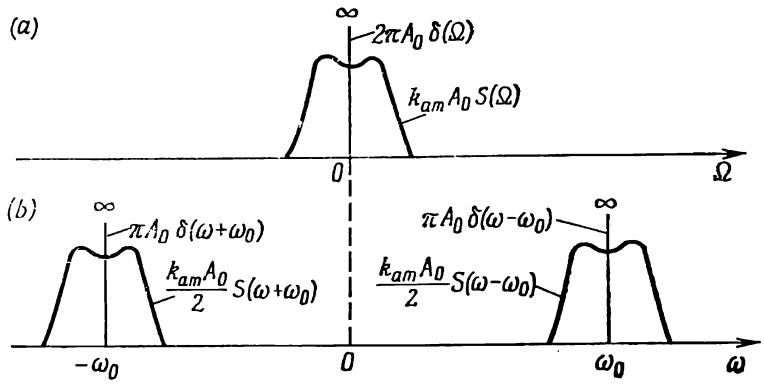


Fig. 3.9. Spectral densities of (*a*) envelope and (*b*) amplitude-modulated oscillation

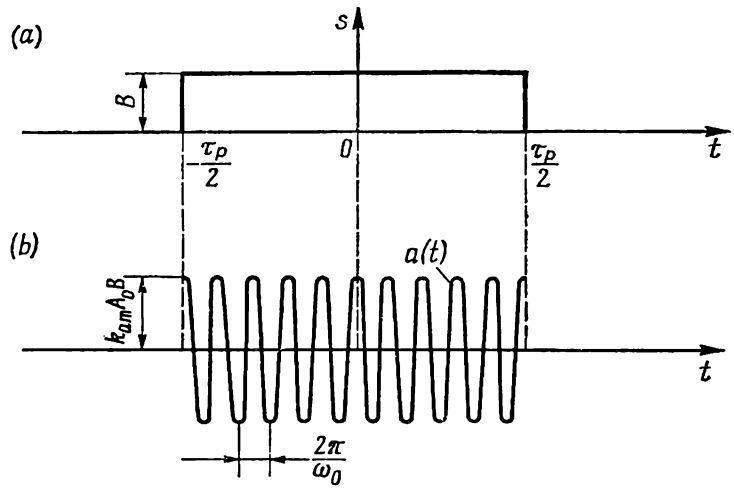


Fig. 3.10. (*a*) Square pulse and (*b*) the same pulse with a radio-frequency carrier ω_0

The spectral density $S_a(\omega)$ of the modulated wave $a(t)$ is shown in Fig. 3.9*b*. In this case the discrete components $\pi A_0\delta(\omega \pm \omega_0)$ represent the carrier oscillation $A_0 \cos(\omega_0 t + \theta_0)$, and the continuous ones, the side frequencies.

If a radio signal does not contain a carrier oscillation of finite amplitude, e.g., when transmitting a solitary radio pulse, there is no discrete component in the spectrum.

Let us consider the spectrum of a square radio pulse (Fig. 3.10*b*) determined by the expression

$$a(t) = \begin{cases} k_{am}A_0B\cos\omega_0 t & \text{for } -\tau_p/2 < t < \tau_p/2 \\ 0 & \text{for } t < -\tau_p/2 \text{ and } t > \tau_p/2 \end{cases} \qquad (3.12)$$

In this example, by the message $s(t)$ is meant a video pulse (Fig. 3.10*a*). The spectral density of such a message [see (2.68)] is given by

$$\mathbf{S}(\Omega) = B \frac{\sin(\Omega\tau_p/2)}{\Omega/2} \tag{3.13}$$

The envelope of the oscillation $a(t)$ is

$$A(t) = k_{am}A_0 s(t)$$

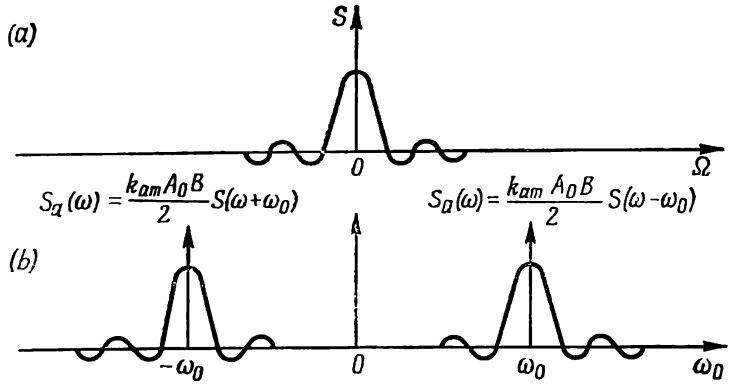


Fig. 3.11. Spectral densities of the functions shown in Fig. 3.10

and the spectral density of this envelope is

$$\mathbf{S}_A(\Omega) = k_{am}A_0\mathbf{S}(\Omega) = k_{am}A_0B \frac{(\sin\Omega\tau_p/2)}{\Omega/2}$$

Since in the given case $\theta_0 = 0$ (Fig. 3.10*b*), formula (3.9) gives

$$\mathbf{S}_a(\omega) = \frac{k_{am}A_0B}{2}\left[\frac{\sin\frac{(\omega-\omega_0)\tau_p}{2}}{\frac{\omega-\omega_0}{2}} + \frac{\sin\frac{(\omega+\omega_0)\tau_p}{2}}{\frac{\omega+\omega_0}{2}}\right] \tag{3.14}$$

The graphs of the spectral densities of the modulating function $s(t)$ and the radio pulse $a(t)$ are shown in Fig. 3.11*a* and *b*.

3.4. ANGLE MODULATION. PHASE AND INSTANTANEOUS FREQUENCY OF A WAVE

In the case of a simple harmonic wave

$$a(t) = A_0 \cos(\omega_0 t + \theta_0) = A_0 \cos\psi(t)$$

the phase advance during some finite interval of time from $t = t_1$ to $t = t_2$ will be

$$\psi(t_2) - \psi(t_1) = (\omega_0 t_2 + \theta_0) - (\omega_0 t_1 + \theta_0) = \omega_0(t_2 - t_1) \tag{3.15}$$

It follows that at a constant angular frequency the phase advance during some interval of time is proportional to the length of this interval.

On the other hand, if it is known that the phase advance during time $t_2 - t_1$ is equal to $\psi(t_2) - \psi(t_1)$, the angular frequency can be determined as the ratio

$$\omega_0 = [\psi(t_2) - \psi(t_1)]/(t_2 - t_1) \tag{3.16}$$

provided, of course, that during this time interval the frequency is constant.

From expression (3.16) it is clear that the angular frequency is nothing else but *the rate of change of the wave phase.*

We now turn our attention to a compound wave in which the frequency can vary in time. In this case equalities (3.15) and (3.16) must be replaced by the following integral and differential expressions:

$$\psi(t_2) - \psi(t_1) = \int_{t_1}^{t_2} \omega(t)\, dt \tag{3.17}$$

$$\omega(t) = \frac{d\psi(t)}{dt} \tag{3.18}$$

In these expressions $\omega(t) = 2\pi f(t)$ is the *instantaneous angular frequency* of the wave and $f(t)$ is the instantaneous frequency expressed in hertzes.

According to expressions (3.17) and (3.18), the phase of a radio-frequency wave at an instant t may be defined as

$$\psi(t) = \int_0^t \omega(t)\, dt + \theta_0 \tag{3.19}$$

where the first term on the right-hand side defines the phase advance during a time interval from zero to t, and θ_0 is the epoch angle of the wave (i.e., the phase angle at the instant $t = 0$).

With such an approach, the phase $\psi(t) = \omega_0 t + \theta(t)$ entering into expression (3.1) must be replaced by

$$\psi(t) = \omega_0 t + \theta(t) + \theta_0 \tag{3.20}$$

Thus, the general expression for a radio-frequency wave having a constant amplitude [i.e., $A(t) = A_0$] and a modulated argument $\psi(t)$ may be written in the form

$$a(t) = A_0 \cos[\omega_0 t + \theta(t) + \theta_0]$$

Expressions (3.18) and (3.19), which establish the relation between the changes in frequency and phase, reveal the common character of the two types of angle modulation—frequency modulation and phase modulation.

Let us clarify relations (3.18) through (3.20) by considering as an example simple harmonic frequency modulation in which the instantaneous frequency of a sine-wave carrier is defined by the expression

$$\omega(t) = \omega_0 + \omega_d \cos \Omega t \tag{3.21}$$

where $\omega_d = 2\pi f_d$ is the amplitude of frequency departure. For simplicity, ω_d will further be called *frequency deviation*, or simply, *deviation*. As in the case of amplitude modulation, ω_0 and Ω denote the carrier frequency and the modulating frequency, respectively.

Let us set up an expression for the instantaneous value of an oscillation (current or voltage) having a constant amplitude and a frequency varying in accordance with expression (3.21).

Substituting $\omega(t)$ from equation (3.21) into (3.19), we obtain

$$\psi(t) = \int_0^t (\omega_0 + \omega_d \cos \Omega t)\, dt + \theta_0$$

Integrating this expression yields

$$\psi(t) = \omega_0 t + (\omega_d/\Omega) \sin \Omega t + \theta_0 \tag{3.22}$$

Thus,

$$a(t) = A_0 \cos [\omega_0 t + (\omega_d/\Omega) \sin \Omega t + \theta_0] \tag{3.23}$$

The phase of the oscillation $a(t)$, along with the linearly increasing term $\omega_0 t$, contains a periodic term $(\omega_d/\Omega) \sin \Omega t$. This makes it possible to regard $a(t)$ as a wave *modulated in phase*. The law governing this modulation is an integral one with respect to the original frequency modulation. It is the frequency modulation according to $\omega_d \cos \Omega t$ that results in phase modulation in accordance with $(\omega_d/\Omega) \sin \Omega t$.

The amplitude of phase variation

$$\theta_{max} = \omega_d/\Omega = m \tag{3.24}$$

is called *angle modulation index*.

It should be noted that the modulation index is completely independent of the centre (unmodulated) frequency ω_0, but is determined exclusively by the deviation ω_d and the modulating frequency Ω.

Now let us consider the opposite case where a wave of stable frequency and phase is made to pass through a device performing a periodic phase modulation according to $\theta(t) = \theta_{max} \sin \Omega t$, so that the wave at the output of the device has the form

$$a\ t) = A_0 \cos [\omega_0 t + \theta_{max} \sin \Omega t + \theta_0] \tag{3.23'}$$

What is the frequency of this output wave? Using expression (3.18), we find

$$\omega(t) = \frac{d}{dt}(\omega_0 t + \theta_{max} \sin \Omega t + \theta_0) = \omega_0 + \theta_{max}\,\Omega \cos \Omega t \qquad (3.21')$$

Taking into account relation (3.24), we arrive at a conclusion that $\theta_{max}\,\Omega = \omega_d$. Thus, a harmonic phase modulation with an

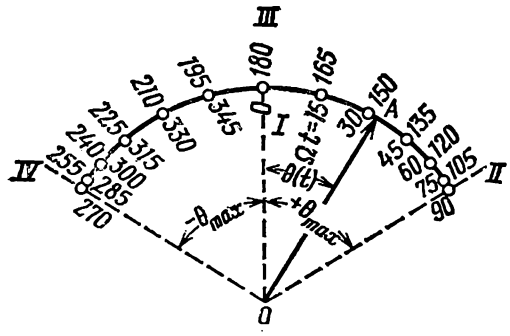


Fig. 3.12. Representation of an angle-modulated radio-frequency oscillation in the form of a rocking vector

index θ_{max} is equivalent to frequency modulation with a deviation $\omega_d = \theta_{max}\,\Omega$.

The above example shows that in the case of a harmonic angle modulation the knowledge of the character of the wave is not sufficient to distinguish whether we deal with frequency or phase modulation. In both cases the vector OA (representing the modulated wave on the diagram of Fig. 3.12) oscillates relative to its initial position in such a manner that the angle θ changes in time as $\theta = \theta_{max} \sin \Omega t$ in phase modulation, and as $\theta = (\omega_d/\Omega) \sin \Omega t = \theta_{max} \sin \Omega t$ in frequency modulation (in which case $\Delta\omega = \omega_d \cos \Omega t$).

The numerals *I*, *II*, *III* and *IV* indicate the position of the vector OA at $\Omega t = 0$, $\pi/2$, π, and $3\pi/2$.

The difference between the phase and frequency modulation manifests itself when the *modulation frequency* Ω *varies.*

In frequency modulation, the *deviation* ω_d *is proportional to the amplitude of the modulating voltage and is independent of the modulation frequency* Ω.

In phase modulation, the *modulation index* $\theta_{max} = m$ *is proportional to the amplitude of the modulating voltage and is independent of the modulation frequency* F.

These statements are clarified by Figs. 3.13 and 3.14 which show the frequency characteristics of f_d and m in frequency and phase modulation. In both cases it is assumed that the modulator input is fed with a modulating voltage of constant amplitude V, whose frequency varies from F_{min} to F_{max}.

In the first case, i.e., in frequency modulation, f_d, which depends only on the amplitude V, is a constant, while the modulation index $m = f_d/F$ diminishes as the modulation frequency increases (Fig. 3.13). In the second case, i.e. in phase modulation, m is independent of F, while $f_d = mF$ changes in proportion to the modulation frequency (Fig. 3.14).

If a *compound* and not a *harmonic* voltage is fed to the modulator input, the structure of the modulated wave in frequency modula-

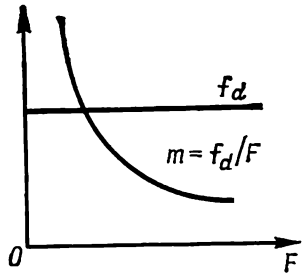


Fig. 3.13. Frequency deviation f_d and modulation index m as functions of the modulation frequency in FM

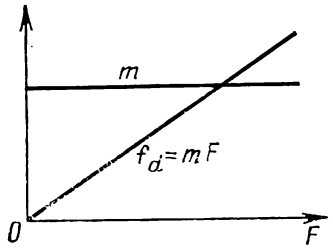


Fig. 3.14. Modulation index m and frequency deviation f_d as functions of the modulation frequency in PM

tion will be *different* from that in phase modulation. In the first case, slow variations of the signal, i.e., low modulation frequencies, correspond to very high values of m (Fig. 3.13), while in the second case, i.e., in phase modulation, they correspond to very low values of f_d (Fig. 3.14).

By way of illustration, let a frequency and a phase modulator be supplied with one and the same sine-wave voltage whose frequency varies from $F_{min} = 200$ Hz to $F_{max} = 2\,000$ Hz. Let us set $f_d = 20$ kHz in frequency modulation (FM) and $\theta_{max} = 0.5$ rad in phase modulation (PM), these values being invariable within the range from 200 to 2 000 Hz at a specified constant amplitude V. In this case, the maximum value of m at F_{min} in FM will be

$$m_{max} = f_d/F_{min} = 20\,000/200 = 100 \text{ rad}$$

while the minimum value of m at F_{max} will be

$$m_{min} = f_d/F_{max} = 10 \text{ rad}$$

In PM, the minimum frequency deviation $f_{d,\,min} = mF_{min} = 100$ Hz will be at the lower modulation frequency F_{min}, while the maximum frequency deviation $f_{d,\,max} = mF_{max} = 1\,000$ Hz will be at the upper modulation frequency F_{max}.

Besides the difference in the wave structure (in modulation by a compound signal), frequency and phase modulation also differ in the

methods of realization. In the first case, the oscillator frequency is directly affected, while in the second, the oscillator produces a stable frequency and the wave is then modulated in phase in the subsequent stages of the system.

3.5. FREQUENCY SPECTRUM OF AN ANGLE-MODULATED WAVE. GENERAL RELATIONS

Let us have a wave

$$a(t) = A_0 \cos [\omega_0 t + \theta(t)] \tag{3.25}$$

of which it is known that the message $s(t)$ being transmitted is contained in the function $\theta(t)$. If the wave $a(t)$ is modulated in phase, functions $\theta(t)$ and $s(t)$ are identical in shape and differ only in the constant coefficient. In this case the spectra of the functions $\theta(t)$ and $s(t)$ apparently coincide with an accuracy to the constant coefficient. In the case of frequency modulation, the function $\theta(t)$ is the integral of the message $s(t)$ being transmitted. This follows from expressions (3.19) and (3.20). Since integration is a linear operation, in the case of frequency modulation the spectrum of the function $\theta(t)$ consists of the same components as the spectrum of the function $s(t)$, but with changed amplitudes and phases.

Disregarding the type of angle modulation (frequency or phase) and assuming that the spectrum of the function $\theta(t)$ is *completely known*, let us find the spectrum of the modulated wave $a(t)$. For this purpose we shall transform expression (3.25) into

$$a(t) = A_0 \cos \theta(t) \cos \omega_0 t - A_0 \sin \theta(t) \sin \omega_0 t$$

$$= a_c(t) - a_s(t) \tag{3.26}$$

From (3.26) it follows that an angle-modulated wave may be regarded as a sum of two quadrature components: a cosine quadrature component $a_c(t) = A_0 \cos \theta(t) \cos \omega_0 t$ and a sine quadrature component $a_s(t) = A_0 \sin \theta(t) \sin \omega_0 t$, each of which is modulated solely in amplitude. For the cosine component the law of amplitude modulation is defined by the slowly varying function $\cos \theta(t)$, and for the sine component, it is defined by the function $\sin \theta(t)$. But in Sec. 3.3 it was found that in order to determine the spectrum of an amplitude-modulated wave, it is sufficient to shift the spectrum of the envelope by an amount ω_0. Therefore, to find the spectrum of the wave $a(t)$ defined by expression (3.26), it is necessary first to find the spectra of the functions $\cos \theta(t)$ and $\sin \theta(t)$, i.e., the spectra of the envelopes of the quadrature components. The shift of these spectra by ω_0 can then be effected by the same method as in the case of ordinary amplitude modulation.

From the above arguments it follows that the transmitted message being the same, the spectrum of an angle-modulated wave is much more complex than that of an amplitude-modulated wave. Indeed,

since cos θ (*t*) and sin θ (*t*) are nonlinear functions of the argument θ (*t*), the spectra of these functions can substantially differ from the spectrum of the function θ (*t*): multiple and combination frequencies may occur, as is the case with common nonlinear transformations of spectrum.

This fact, as well as the presence of two quadrature components, suggests that in the case of angle modulation the spectrum of the modulated wave cannot be obtained by simply shifting the message spectrum by an amount equal to the carrier frequency ω_0, as is the case with amplitude modulation. The relation between the spectrum of the message and that of the modulated wave in angle modulation is more complex.

3.6. WAVE SPECTRUM IN HARMONIC ANGLE MODULATION

Let us use the above results to analyze a wave of the form

$$a(t) = A_0 \cos(\omega_0 t + m \sin \Omega t) \qquad (3.25')$$

This expression agrees with (3.23) and (2.23′) when the modulating function is $\omega(t) = \omega_0 + \omega_d \cos \Omega t$. The epoch angle θ_0, as well as the epoch angle γ of the modulating function, is omitted for convenience in subsequent work. If necessary, these can easily be introduced into final expressions.

In our case, $\theta(t) = m \sin \Omega t$. Substituting θ (*t*) into expression (3.26), we obtain

$$a(t) = A_0 \cos(m \sin \Omega t) \cos \omega_0 t - A_0 \sin(m \sin \Omega t) \sin \omega_0 t \qquad (3.27)$$

Since the factors cos (*m* sin Ω*t*) and sin (*m* sin Ω*t*) are periodic functions of time, we can expand them in a cosine and a sine Fourier series.

In the theory of Bessel functions the following relations are derived:

$$\sin(m \sin \Omega t) = 2J_1(m) \sin \Omega t + 2J_3(m) \sin 3\Omega t + 2J_5(m) \sin 5\Omega t + \ldots \qquad (3.28)$$

$$\cos(m \sin \Omega t) = J_0(m) + 2J_2(m) \cos 2\Omega t + 2J_4(m) \cos 4\Omega t + \ldots \qquad (3.29)$$

$$\sin(m \cos \Omega t) = 2J_1(m) \cos \Omega t - 2J_3(m) \cos 3\Omega t + 2J_5(m) \cos 5\Omega t - \ldots \qquad (3.28')$$

$$\cos(m \cos \Omega t) = J_0(m) - 2J_2(m) \cos 2\Omega t + 2J_4(m) \cos 4\Omega t - \ldots \qquad (3.29')$$

Here $J_n(m)$ is a Bessel function of the first kind and *n*th order with argument *m*.

Using relations (3.28) and (3.29), equation (3.27) may be written as

$$a(t) = A_0 [J_0(m) \cos \omega_0 t - 2J_1(m) \sin \Omega t \sin \omega_0 t + 2J_2(m) \cos 2\Omega t \cos \omega_0 t - 2J_3(m) \sin 3\Omega t \sin \omega_0 t + \ldots] \quad (3.30)$$

or, in a more expanded form,

$$a(t) = A_0 \cos(\omega_0 t + m \sin \Omega t) = A_0 \{J_0(m) \cos \omega_0 t + J_1(m) [\cos(\omega_0 + \Omega) t - \cos(\omega_0 - \Omega) t] + J_2(m) [\cos(\omega_0 + 2\Omega) t + \cos(\omega_0 - 2\Omega) t] + J_3(m) [\cos(\omega_0 + 3\Omega) t - \cos(\omega_0 - 3\Omega) t] + \ldots\} \quad (3.31)$$

Thus, in frequency and phase modulation, the modulated wave spectrum consists of an infinite number of pairs of side frequencies, arranged symmetrically about the carrier frequency ω_0 and differing from it by $n\Omega$, where n is any integer. The amplitude of the nth sideband component is $A_n = J_n(m) A_0$, where A_0 is the amplitude of the unmodulated wave and m is the modulation index. From this it follows that the relative contribution of the various sideband components to the total power of the resultant wave is determined by the value of m.

Let us consider angle-modulation conditions for low and high values of m. When $m \ll 1$, we may safely write

$$\sin(m \sin \Omega t) \approx m \sin \Omega t, \cos(m \sin \Omega t) \approx 1,$$

and expression (3.27) then becomes

$$a(t) \approx A_0 (\cos \omega_0 t - m \sin \Omega t \sin \omega_0 t) = A_0 \left[\cos \omega_0 t + \frac{m}{2} \cos(\omega_0 + \Omega) t - \frac{m}{2} \cos(\omega_0 - \Omega) t\right] \quad (3.32)$$

Let us compare this expression with the equation for an amplitude-modulated wave having the same modulating function (i.e., the information to be transmitted) as in frequency modulation. Since the expression (3.32) is obtained from (3.25′) for frequency modulation according to $\omega(t) = \omega_0 + \omega_d \cos \Omega t$, then, for convenience in comparison, we shall set the modulating function for the amplitude-modulated wave to have a similar form $A(t) = A_0 (1 + M \cos \Omega t)$. In this case the amplitude-modulated wave will be given by

$$a_{am}(t) = A_0 (1 + M \cos \Omega t) \cos \omega_0 t = A_0 \left[\cos \omega_0 t + \frac{M}{2} \cos(\omega_0 + \Omega) t + \frac{M}{2} \cos(\omega_0 - \Omega) t\right] \quad (3.33)$$

From comparison of (3.32) with (3.33) it is clear that for low values of m, the wave spectrum, as in the case of amplitude modulation, consists of the carrier frequency ω_0 and two side frequencies —

upper, $\omega_0 + \Omega$, and lower, $\omega_0 - \Omega$, the only difference being in their phasing with respect to the carrier wave. In amplitude modulation, the phases of the sideband waves are symmetrical with respect to the carrier frequency, while in angle modulation, the phase of the lower sideband wave is shifted through 180° [the minus sign before the last term in (3.32)]. This is illustrated by the vector dia-

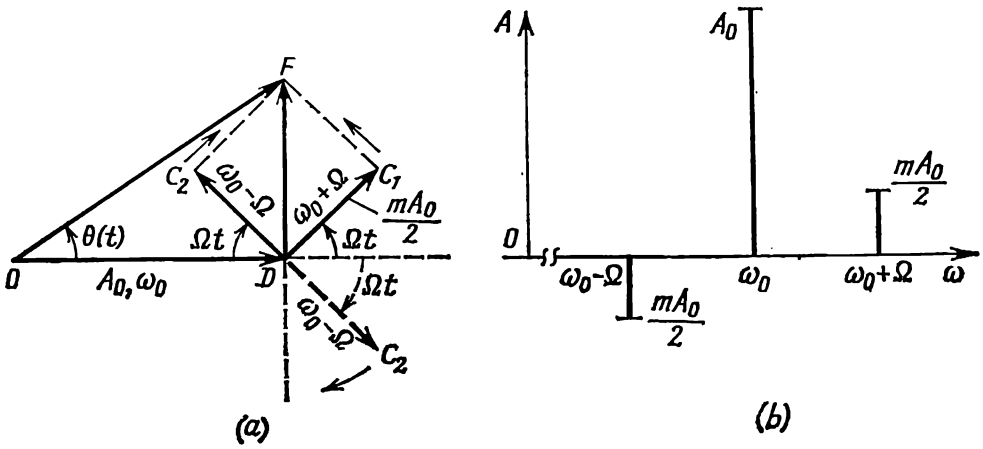


Fig. 3.15. (*a*) Vector diagram and (*b*) oscillation spectrum in angle modulation with $m \ll 1$

gram shown in Fig. 3.15*a*. The direction of the vector DC_2 in amplitude modulation is shown by a dotted line. The 180-degree change in the direction of this vector results in that the modulation vector DF in angle modulation is always normal to the direction of the vector OD representing the carrier wave. The vector OF representing the resultant wave changes both in phase and in amplitude. However, when $m = \theta_{max} \ll 1$, the changes in amplitude are so small that they may be neglected, and to a first approximation, modulation in this case may be regarded as purely phase modulation.

A spectral diagram for the case of angle modulation with $m \ll 1$ is shown in Fig. 3.15*b*. Here, the equality of the amplitudes of the side waves is retained, but the phase of the lower side-frequency wave is shifted through 180°. The amplitudes of the side waves are equal to $mA_0/2$, therefore, in the given case, the modulation index m is equal in magnitude to the factor M characterizing the degree of amplitude variation in amplitude modulation. It should be noted that for $m \ll 1$, the spectrum width is equal to 2Ω, as in the case of amplitude modulation. This result shows that for very small deviations ω_d (as compared with Ω) the width of the spectrum is independent of ω_d.

As the phase deviation increases, i.e., as m is increased, equation (3.32) and the diagram in Fig. 3.15*a* can no longer be used to illu-

strate exactly the phenomena taking place in the process of phase or frequency modulation. This stems from the fact that an oscillation of strictly constant amplitude, whose frequency or phase varies within a wide range cannot be represented in terms of a carrier and a single pair of side frequencies. To obtain a correct picture, it is necessary to take into account the side frequencies of higher orders, in accordance with expression (3.31).

For m ranging between 0.5 and 1, the second pair of side frequencies becomes of some importance, and, the width of the spectrum

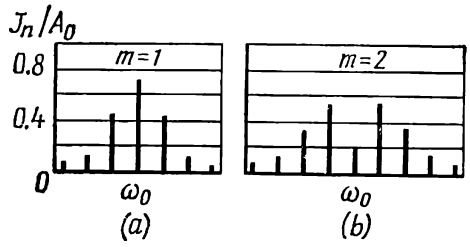


Fig. 3.16. Spectra of an angle-modulated oscillation
(a) for $m = 1$; (b) for $m = 2$

must therefore be taken equal to 4Ω. Then, for $1 < m < 2$, one has to take into account the third and the fourth pair of the side frequencies, and so on. The spectrograms for $m = 1$ and $m = 2$ are given in Fig. 3.16*a* and *b*. In these figures no account is taken of the phases of the oscillations, however, it should be noted that for

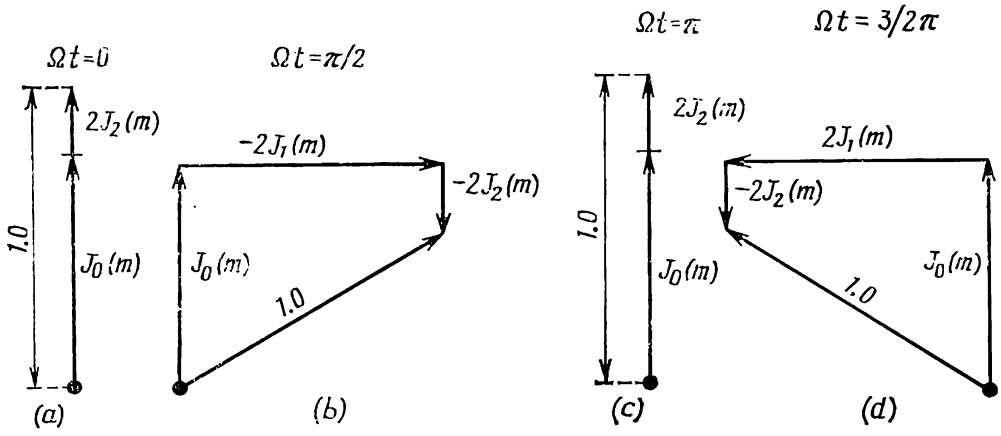


Fig. 3.17. Phasing of side-frequency oscillations at various instants of time

odd n the amplitudes of the lower side-frequency waves should be taken with a minus sign. The amplitudes of all the components of the spectrum are represented by vertical lines having a length $J_n(m)$, and the distances from the line $J_0(m)$ representing the amplitude of

the carrier wave are equal to $n\Omega$, where Ω is the modulation frequency and n is the serial number of the side frequency. The amplitude of the resultant wave is taken at 100%, i. e., $A_0 = 1$; the magnitudes of $J_n(m)$ given in the diagrams indicate the amplitudes of the waves of the corresponding frequencies in terms of fractions of the amplitude of the resultant wave.

The vector diagrams for the instants $\Omega t = 0$, $\pi/2$, π, and $3\pi/2$ and for $m = 1$, plotted in accordance with expression (3.30), are shown in Fig. 3.17*a*, *b*, *c*, and *d*.

Now let us consider the case of large values of m. What we have to do is to determine the dependence of the Bessel function $J_n(m)$ on the serial number n for large values of the argument m. It has been found that when $m \gg 1$, the quantity $|J_n(m)|$ is more or less uniform for all integer values of $|n|$ smaller than the argument m. When $|n|$ is close to m, $|J_n(m)|$ forms a spike, and with further increase in $|n|$ the function $|J_n(m)|$ quickly goes to zero. The general character of this dependence is shown in Fig. 3.18 for $m = 100$. From this figure it is clear that the greatest serial number n of the side frequency that should still be taken account of is approximately equal to the modulation index m (in this case $n = 100$).

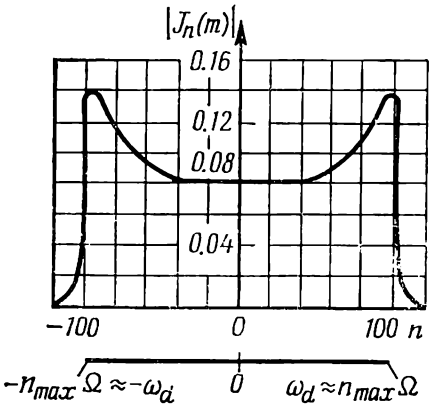


Fig. 3.18. Spectrum width of a frequency-modulated oscillation for high values of the modulation index m

Equating this maximum value n_{max} to m, we arrive at a conclusion that the total spectrum width of the modulated wave is

$$2\,|n_{max}|\,\Omega \approx 2m\Omega$$

But, $m = \omega_d/\Omega$, hence, for *large magnitudes of the modulation index the spectrum width of the modulated wave is close to twice the frequency deviation*:

$$2\,|n_{max}|\,\Omega \approx 2\omega_d \qquad (3.34)$$

It should be noted that, according to the definition of m [see (3.24)], the expression "modulation with a low modulation index" is equivalent to the expression "fast modulation", and the expression "modulation with a high modulation index", to "slow modulation". Therefore, we may state that in *fast angle modulation* (when $\omega_d \ll \Omega$) the width of the spectrum of the modulated wave is close to 2Ω, and that in *slow angle modulation* (when $\omega_d \gg \Omega$), it is close to $2\omega_d$.

3.7. SPECTRUM OF A RADIO PULSE WITH A FREQUENCY-MODULATED CARRIER

When the modulating function in frequency modulation is other than harmonic, the task of finding the wave spectrum becomes more complicated. The choice of a most convenient method of analysis depends on the character of the modulating function. Let us explain one of possible methods by considering as an example a widely used signal—the so-called *linear FM pulse*. Such a signal is shown in Fig. 3.19*a*, and its modulating function is illustrated in Fig. 3.19*b*.

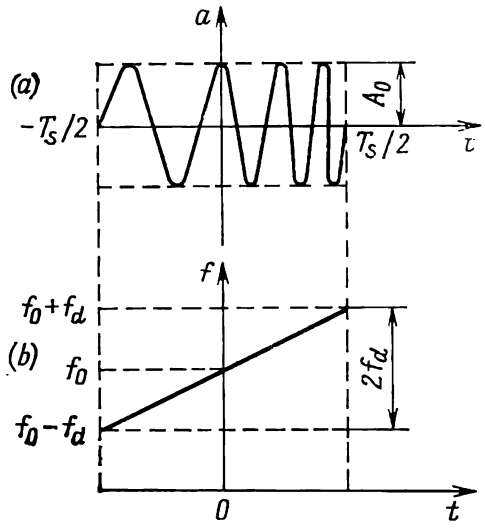


Fig. 3.19. (*a*) Linear FM pulse and (*b*) variation of its carrier frequency

The instantaneous pulse carrier frequency $\omega(t) = 2\pi f(t)$ may be defined by the expression

$$\omega(t) = \omega_0 + \beta t, \quad |t| \leqslant T_s/2 \tag{3.35}$$

where

$$\beta = 2\omega_d/T_s = 2 \times 2\pi f_d/T_s \tag{3.36}$$

is the rate of linear frequency change within the pulse. Then, the instantaneous value of the oscillation shown in Fig. 3.19*a* may be written in the form

$$a(t) = A_0 \cos\left(\int \omega(t)\,dt\right) = A_0 \cos\left(\omega_0 t + \frac{\beta t^2}{2}\right), \quad -\frac{T_s}{2} < t < \frac{T_s}{2} \tag{3.37}$$

The product of the full frequency deviation and the pulse duration

$$2f_d T_s = m \tag{3.38}$$

is the main parameter of a linear FM signal. From comparison of expression (3.38) with (2.126) it can be inferred that m may be treated as the base of the linear FM signal.

Taking into account (3.38), expression (3.36) may be written in the form

$$\beta = 2\pi m/T_s^2 \tag{3.39}$$

In this case the signal $a(t)$ is defined by the expression

$$a(t) = A_0 \cos\left(\omega_0 t + \frac{\pi m t^2}{T_s^2}\right), \quad -\frac{T_s}{2} < t < \frac{T_s}{2} \tag{3.40}$$

Let us define the spectral density of this signal by means of the general expression (2.48):

$$S(\omega) = A_0 \int_{-T_s/2}^{T_s/2} \cos\left(\omega_0 t + \frac{\pi m t^2}{T_s^2}\right) e^{-i\omega t}\, dt$$

$$= \frac{A_0}{2} \int_{-T_s/2}^{T_s/2} \exp\left\{i\left[\frac{\pi m t^2}{T_s^2} - (\omega - \omega_0)\, t\right]\right\} dt$$

$$+ \frac{A_0}{2} \int_{-T_s/2}^{T_s/2} \exp\left\{-i\left[\frac{\pi m t^2}{T_s^2} + (\omega + \omega_0)\, t\right]\right\} dt \qquad (3.41)$$

The first term on the right-hand side of the above expression determines the spike of the spectral density near the frequency $\omega = \omega_0$, while the second term, the other spike near the frequency $\omega = -\omega_0$.

When determining $S(\omega)$ in the positive frequency region, the second term may be omitted [see formula (3.10)]. In the first term, it is advisable to complete the square in the exponent in the integrand (β is considered to be positive)

$$\frac{\pi m t^2}{T_s^2} - (\omega - \omega_0)\, t = \left[\frac{\pi m t^2}{T_s^2} - (\omega - \omega_0)\, t + d^2\right] - d^2$$

$$= \left(\frac{\sqrt{\pi m}\, t}{T_s} - d\right)^2 - d^2 \qquad (3.42)$$

where

$$d = (\omega - \omega_0)\, T_s / 2\sqrt{\pi m} \qquad (3.43)$$

Substituting expression (3.42) into (3.41) and transforming to the new variable

$$y = \sqrt{\pi m}\, t / T_s - d \qquad (3.44)$$

we obtain

$$S(\omega) = \frac{A_0}{2} e^{-id^2} \int_{-T_s/2}^{T_s/2} \exp\left[i\left(\frac{\sqrt{\pi m}\, t}{T_s} - d\right)^2\right] dt = \frac{A_0}{2} \frac{T_s}{\sqrt{\pi m}}$$

$$\times \exp\left[-i\frac{(\omega - \omega_0)^2}{2\beta}\right] \int_{-\sqrt{\frac{\pi m}{4}}\left(1 + \frac{\omega - \omega_0}{\omega_d}\right)}^{\sqrt{\frac{\pi m}{4}}\left(1 - \frac{\omega - \omega_0}{\omega_d}\right)} e^{-iy^2}\, dy \qquad (3.45)$$

Let us designate

$$u_1 = \sqrt{\frac{\pi m}{4}}\left(1 - \frac{\omega - \omega_0}{\omega_d}\right) \qquad (3.46)$$

and

$$u_2 = \sqrt{\frac{\pi m}{4}}\left(1 + \frac{\omega - \omega_0}{\omega_d}\right) \tag{3.47}$$

In these terms the expression for S (ω) assumes the form

$$\mathbf{S}(\omega) = \frac{A_0 T_s}{2\sqrt{\pi m}} \exp\left[-i\frac{(\omega-\omega_0)^2}{2\beta}\right]\left[\int_0^{u_1} e^{iy^2}\,dy + \int_0^{u_2} e^{iy^2}\,dy\right] \tag{3.48}$$

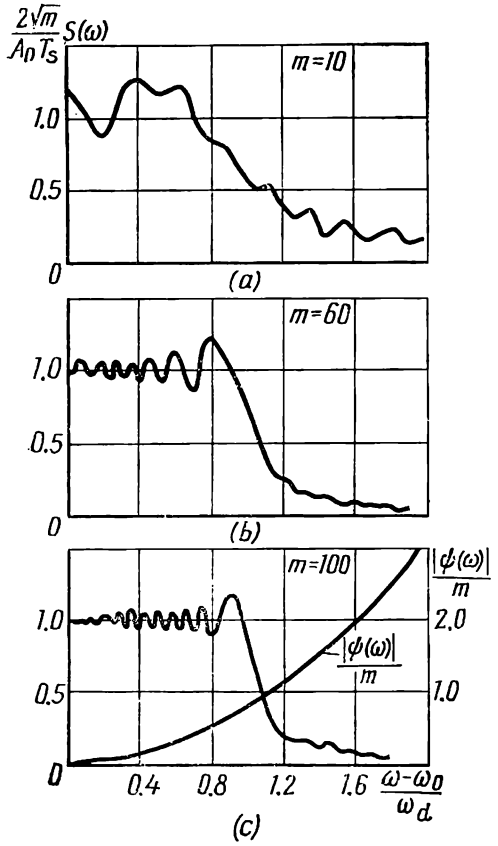


Fig. 3.20. Spectral density of a linear FM pulse for various values of the base $m = 2f_d T_s$

Recall the relation

$$\sqrt{\frac{2}{\pi}}\int_0^x e^{iy^2}\,dy = \sqrt{\frac{2}{\pi}}\int_0^x \cos y^2\,dy + i\sqrt{\frac{2}{\pi}}\int_0^x \sin y^2\,dy = C(x) + iS(x)$$

where $C(x)$ and $S(x)$ are Fresnel integrals.

Using this relation, expression (3.48) can be reduced to the form

$$\mathbf{S}(\omega) = \frac{A_0 T_s}{2\sqrt{\pi m}}\sqrt{\frac{\pi}{2}} \times \exp\left[-\frac{i(\omega-\omega_0)^2}{2\beta}\right] \times \{C(u_1) + C(u_2) + i[S(u_1) + S(u_2)]\} \tag{3.49}$$

From (3.49) it follows that the modulus of the spectral density of the signal in question is

$$S(\omega) = \frac{T_s A_0}{2\sqrt{m}}\frac{1}{\sqrt{2}}\sqrt{[C(u_1)+C(u_2)]^2 + [S(u_1)+S(u_2)]^2} \tag{3.50}$$

while the phase characteristic of the spectrum for $\omega > 0$ is given by

$$\psi(\omega) = -\frac{(\omega-\omega_0)^2}{2} + \arctan\frac{S(u_1)+S(u_2)}{C(u_1)+C(u_2)} = -\frac{\pi m}{4}\frac{(\omega-\omega_0)^2}{\omega_d^2} + \arctan\frac{S(u_1)+S(u_2)}{C(u_1)+C(u_2)} \tag{3.51}$$

The diagrams of $(2\sqrt{m}/A_0T_s)\,S(\omega)$ versus $(\omega-\omega_0)/\omega_d$ (Fig. 3.20 *a*, *b* and *c*) show that for high m the shape of $S(\omega)$ approaches a rectangle and the spectrum width is close to $2\omega_d$. In this case the phase characteristic assumes the form of a quadratic parabola (Fig. 3.20*c*). The second term in (3.51), which tends to the constant $\pi/4$, is omitted.

When $\omega=\omega_0$, $u_1=u_2=\sqrt{\pi m/4}$, so that for high m and $\omega=\omega_0$, when $C(u_1)\approx C(u_2)\approx 0.5$ and $S(u_1)\approx S(u_2)\approx 0.5$, the square root in expression (3.50) becomes $\sqrt{2}$, and $S(\omega)\to A_0T_s/2\sqrt{m}$.

3.8. SPECTRUM OF AN AMPLITUDE-FREQUENCY MODULATED WAVE

Let us generalize expressions (3.25) and (3.26) by replacing therein the constant amplitude A_0 by one, $A(t)$, which is a function of time:

$$a(t)=A(t)\cos[\omega_0t+\theta(t)]$$
$$=A(t)\cos\theta(t)\cos\omega_0t-A(t)\sin\theta(t)\sin\omega_0(t)$$
$$=a_c(t)-a_s(t) \tag{3.52}$$

As in Sections 3.5 and 3.6, determining the spectrum is reduced to finding the spectra of the functions $A_c(t)=A(t)\cos\theta(t)$ and $A_s(t)=A(t)\sin\theta(t)$, i.e., of the envelopes of the quadrature components, and then shifting these spectra by an amount ω_0.

Let us designate the spectral densities of the functions $A_c(t)$ and $A_s(t)$ by the symbols $\mathbf{S}_{Ac}(\omega)$ and $\mathbf{S}_{As}(\omega)$. Then

$$\left.\begin{aligned}\mathbf{S}_{Ac}(\omega)&=\int_{-\infty}^{\infty}A_c(t)\,e^{-i\omega t}\,dt=\int_{-\infty}^{\infty}A(t)\cos\theta(t)\,e^{-i\omega t}\,dt\\ \text{and}\\ \mathbf{S}_{As}(\omega)&=\int_{-\infty}^{\infty}A_s(t)\,e^{-i\omega t}\,dt=\int_{-\infty}^{\infty}A(t)\sin\theta(t)\,e^{-i\omega t}\,dt\end{aligned}\right\} \tag{3.53}$$

The spectral density of the cosine quadrature component $a_c(t)=A_c(t)\cos\omega_0t$, in accordance with expression (2.58) (at $\theta_0=0$), is given by

$$\mathbf{S}_{ac}(\omega)=\frac{1}{2}[\mathbf{S}_{Ac}(\omega-\omega_0)+\mathbf{S}_{Ac}(\omega+\omega_0)] \tag{3.54}$$

When determining the spectrum of the sine quadrature component $a_s(t)=A_s(t)\sin\omega_0(t)$, the phase angle θ_0 in (2.58) should be taken equal to $-90°$. Consequently,

$$\mathbf{S}_{as}(\omega)=-\frac{i}{2}[\mathbf{S}_{As}(\omega-\omega_0)+\mathbf{S}_{As}(\omega+\omega_0)] \tag{3.54'}$$

In the positive frequency region we may assume that

$$S_{Ac}(\omega + \omega_0) \approx 0, \quad S_{As}(\omega + \omega_0) \approx 0$$

Thus, finally, the spectral density of the oscillation $a(t) = a_c(t) - a_s(t)$ is defined by the expression

$$S_a(\omega) = S_{ac}(\omega) - S_{as}(\omega) = \frac{1}{2}[S_{Ac}(\omega - \omega_0) + iS_{As}(\omega - \omega_0)], \quad \omega > 0 \tag{3.55}$$

Transforming to the new variable $\Omega = \omega - \omega_0$, we obtain

$$S_a(\omega_0 + \Omega) = \frac{1}{2}[S_{Ac}(\Omega) + iS_{As}(\Omega)] \tag{3.56}$$

The structure of the spectrum of the oscillation $a(t)$ in combined amplitude-frequency modulation depends on the form of the functions $A(t)$ and $\theta(t)$ and the relation between them.

In purely amplitude modulation, the spectrum of the oscillation $a(t)$ is characterized by the absolute symmetry of the amplitudes and phases of the side waves relative to the carrier; in purely angle modulation [$A(t) = A_0 = \text{const}$], the phases of the lower side-frequency waves for odd n are shifted through 180° with respect to the upper side-frequency waves (see Sec. 3.6). Simultaneous modulation of a wave in both amplitude and angle may, for some relations between $A(t)$ and $\theta(t)$, lead to the asymmetry of the spectrum $S_a(\omega_0 + \Omega)$ with respect to ω_0 not only in phase but also in amplitude. In particular, if $\theta(t)$ is an odd function of t, the spectrum of the oscillation $a(t)$ is asymmetrical, no matter what the function $A(t)$ is.

Indeed, let $A(t)$ be an even function. Then the product $A(t)\cos\theta(t) = A_c(t)$ is an even function of t, while $A(t)\sin\theta(t) = A_s(t)$ is an odd function of t, and, in accordance with the properties of the Fourier transform, described in Sec. 2.7-6, $S_{Ac}(\Omega)$ is a real and even function of Ω, while $S_{Ac}(\Omega)$ is an imaginary and odd one. With account being taken of the factor i, the second term in (3.56) is also a real but odd function of Ω and, therefore, the spectral density $S_a(\omega)$ is a real function which is asymmetrical relative to the point $\omega = \omega_0$. An example of such a spectrum is given in Fig. 3.21. (With respect to the point $\omega = 0$ the modulus of the spectral density is symmetrical under any conditions.)

A similar result is also obtained for an odd function $A(t)$. In this case, $S_{Ac}(\Omega)$ is an odd, imaginary function of Ω, while $S_{As}(\Omega)$ is an even and real function. The term $iS_{As}(\Omega)$ in expression (3.56) becomes imaginary, and the modulus of the sum $S_{Ac}(\Omega) + iS_{As}(\Omega)$ then becomes an asymmetrical function with respect to the point $\omega = \omega_0$.

By similar reasoning, we may easily show that for the spectrum $S_a(\omega)$ to be symmetrical it is required that $\theta(t)$ should be an even function and that the function $A(t)$ should simultaneously be either even or odd function of t. If the function $A(t)$ can be represented as a sum of an even and an odd component, the spectrum $S_a(\omega)$ is asymmetrical even when the function $\theta(t)$ is even. For example, the linear FM pulse discussed in Sec. 3.7 has a symmetrical spectrum. In this case the rectangular envelope is an even function with respect to t, as well as the function $\theta(t) = \frac{1}{2}\beta t^2$, provided the time origin is properly selected.

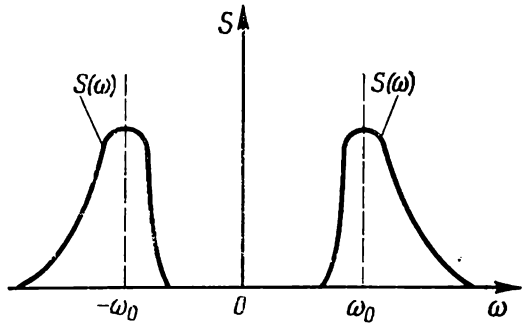


Fig. 3.21. Example of an asymmetric spectrum in combined amplitude and frequency modulation

One can get a clear idea of the deformation of the spectrum of an oscillation subjected to combined amplitude-angle modulation by considering a case where both types of modulation are effected by means of a harmonic modulating function. In order to simplify the analysis, let this function be given in the form of $\cos \Omega t$ for angle modulation, and in the form of $\cos \Omega t$ or $\sin \Omega t$ for amplitude modulation.

1. Both functions — $A(t)$ and $\theta(t)$ — are even with respect to t:

$$A(t) = A_0 (1 + M \cos \Omega t), \quad \theta(t) = m \cos \Omega t,$$
$$M \ll 1, \quad m \ll 1$$

Expression (3.52) assumes the form

$$a(t) = A_0 (1 + M \cos \Omega t) \cos [\omega_0 t + m \cos \Omega t]$$

Assuming, as in Sec. 3.3, that the approximate equalities $\cos(m \cos \Omega t) \approx 1$, $\sin m(\cos \Omega t) \approx m \cos \Omega t$ hold true, let us reduce this expression to a form similar to (3.32):

$$a(t) = A_0 \left[(1 + M \cos \Omega t) \cos \omega_0 t - m \left(\frac{M}{2} + \cos \Omega t + \frac{M}{2} \cos 2\Omega t \right) \sin \omega_0 t \right]$$

$$= A_0 \left\{ \cos \omega_0 t + \frac{M}{2} [\cos(\omega_0 + \Omega) t + \cos(\omega_0 - \Omega) t] - m \left[\frac{M}{2} \sin \omega_0 t + \frac{1}{2} \sin(\omega_0 + \Omega) t + \frac{1}{2} \sin(\omega_0 - \Omega) t \right] - \frac{mM}{4} [\sin(\omega_0 + 2\Omega) t + \sin(\omega_0 - 2\Omega) t] \right\}$$

Adding together the quadrature components $\cos \omega_0 t$ and $(mM/2) \sin \omega_0 t$, we obtain the expression for the amplitude of the resultant oscillation at frequency ω_0: $\sqrt{1 + (mM/2)^2}$ with $A_0 = 1$. In a similar way we find the amplitude $\frac{1}{2}\sqrt{M^2 + m^2}$ for oscillations of frequencies $\omega_0 \pm \Omega$, and $mM/4$ for those of frequencies $\omega_0 \pm 2\Omega$. The spectrum of the oscillation $a(t)$ is shown in Fig. 3.22*a*. The amplitude spectrum is *symmetrical*.

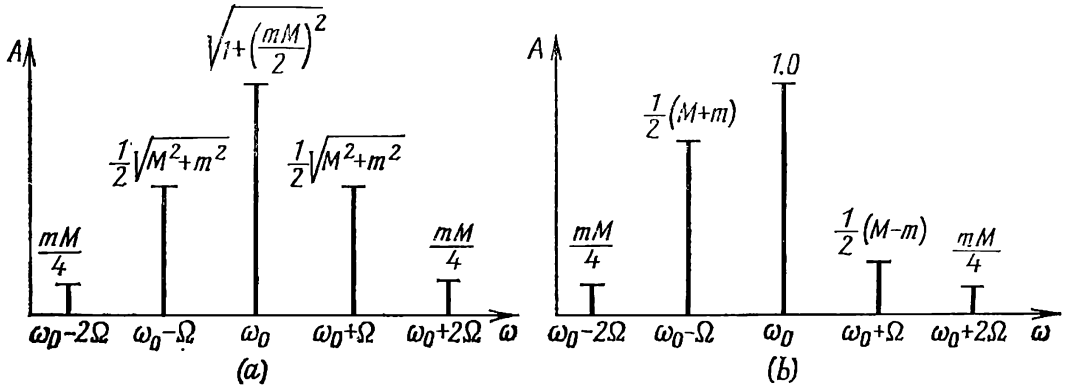


Fig. 3.22. Oscillation spectrum in combined amplitude-frequency modulation by a harmonic function

2. The function $\theta(t)$ is even, while the function $A(t)$ contains an odd component:

$$A(t) = A_0 (1 + M \sin \Omega t)$$
$$\theta(t) = m \cos \Omega(t), \quad M \ll 1, \quad m \ll 1$$

Mathematical manipulations similar to the preceding ones give the following results for the amplitudes of the resultant oscillations: 1 at frequency ω_0, $\frac{1}{2}(M - m)$ at frequency $\omega_0 + \Omega$, $\frac{1}{2}(M + m)$ at frequency $\omega_0 - \Omega$, and $mM/4$ at frequencies $\omega_0 \pm 2\Omega$.

The spectral diagram is shown in Fig. 3.22*b*.

In this example, the symmetry of the spectrum is disturbed because of the difference in the amplitudes of the oscillations of the upper and lower side frequencies $\omega_0 \pm \Omega$.

The disturbance of the spectrum symmetry in combined amplitude-angle modulation may sometimes be used as an indication of incorrect operation of a device intended to perform purely amplitude modulation: a distortion of the spectrum in this case is evidence that the useful amplitude modulation is accompanied by a parasitic angle modulation.

3.9. ENVELOPE, PHASE, AND FREQUENCY OF A NARROW-BAND SIGNAL

The present state of radio engineering is characterized by continuous improvements in the methods of transmitting information. This development follows the line of finding new types of signals and new methods of their treatment.

The modulated waves we have considered in the preceding sections are but the simplest types of radio signals. Radio engineers often have to deal with radio signals which are modulated both in amplitude and in frequency (or phase) according to a fairly complex law.

In any case, the given signal $a\,(t)$ is supposed to be a narrow-band process. This means that all the spectral components of the signal are concentrated within a rather narrow band as compared to some centre frequency ω_0. When such a signal is given in the form

$$a\,(t) = A\,(t)\cos\psi\,(t) \tag{3.57}$$

an ambiguity arises in the selection of the functions $A\,(t)$ and $\psi\,(t)$ since whatever the function $\psi\,(t)$, equation (3.52) can always be satisfied by properly selecting the function $A\,(t)$.

For example, the simple (harmonic) oscillation

$$a\,(t) = A_0\cos\omega_0 t \tag{3.58}$$

can, if necessary, be represented in the form

$$a\,(t) = A\,(t)\cos\omega t \tag{3.58'}$$

where $\omega = \omega_0 + \Delta\omega$.

In expression (3.58') the envelope $A\,(t)$, in contrast to A_0, is a function of time. This function can be determined by the condition

$$A_0\cos\omega_0 t = A(t)\cos(\omega_0 + \Delta\omega)\,t$$

Whence

$$A\,(t) = \frac{A_0\cos\omega_0 t}{\cos(\omega_0+\Delta\omega)\,t} = \frac{A_0\cos\omega_0 t}{\cos\Delta\omega t\cos\omega_0 t - \sin\Delta\omega t\sin\omega_0 t}$$

$$= \frac{A_0}{\cos\Delta\omega t - \sin\Delta\omega t\tan\omega_0 t} \tag{3.59}$$

This example shows that incorrect selection of $\psi(t)$ (ωt instead of $\omega_0 t$) considerably complicates the expression for $A\,(t)$, this new function $A\,(t)$ actually being not an "envelope" in the universally adopted sense, since it can intersect the curve $a\,(t)$ (instead of touching it at points where $a\,(t)$ is maximum). Manipulating such an "envelope" makes no sense and, in some cases, is inadmissible, because this may lead to erroneous practical results (for example, when considering the operation of an amplitude detector).

This trouble can be avoided by representing $A\,(t)$ and $\psi\,(t)$ by the following relations:

$$A\,(t) = \sqrt{a^2(t) + a_1^2\,(t)} \tag{3.60}$$

$$\psi\,(t) = \arctan\,[a_1(t)/a\,(t)] \tag{3.61}$$

where $a_1(t)$ is a new function related to the original function by the relations

$$a_1(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{a(\tau)}{\tau - t}\,d\tau \tag{3.62}$$

$$a(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{a_1(\tau)}{\tau - t}\,d\tau \tag{3.63}$$

These relations are called the Hilbert-transform pair, and the function $a_1(t)$ is the Hilbert transform (harmonic conjugate) of the function $a(t)$.

In accordance with expressions (3.60) and (3.61), Fig. 3.23 shows the function $a(t)$ as the projection of the vector $\mathbf{A}(t)$ on the abscissa axis from which the angle $\psi(t)$ is read.

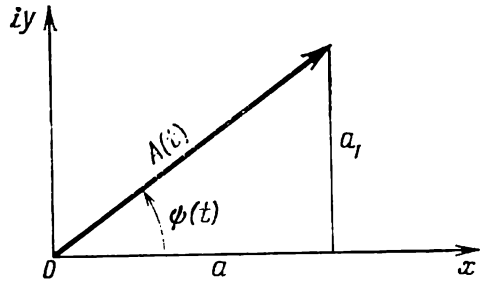


Fig. 3.23. Defining the envelope of a radio-frequency oscillation after Hilbert

To clarify the meaning of expressions (3.60) and (3.61), as well as the sense of the requirement that $a_1(t)$ should be the harmonic conjugate of the original function $a(t)$, let us first consider some properties of $A(t)$ stemming directly from expression (3.60) and holding true for any function $a_1(t)$.

First of all, we can see that at points where $a_1(t) = 0$ the equality $A(t) = a(t)$ holds.

Differentiating (3.60), we get

$$A\frac{dA}{dt} = a\frac{da}{dt} + a_1\frac{da_1}{dt}$$

From this it is obvious that at $a_1 = 0$, when $A(t) = a(t)$, the following additional equality also holds:

$$\frac{dA}{dt} = \frac{da}{dt}$$

Consequently, at points where $a_1(t) = 0$, the curves $A(t)$ and $a(t)$ have common tangents.

These conditions, however, are not sufficient for regarding the curve $A(t)$ as the "simplest" envelope of the rapidly oscillating function $a(t)$. We must also require that the curve $A(t)$ should touch the curve $a(t)$ at points where the latter has the peak value (or one sufficiently close to it). In other words, at points where $a_1(t) = 0$, the function $a(t)$ must have values close to the peak ones.

It is precisely this condition that will be satisfied, provided the function $a_1(t)$ is the Hilbert transform of the function $a(t)$. This property of the Hilbert transforms can be best illustrated, taking a harmonic signal as an example.

Let $a(t) = \cos \omega_0 t$, $-\infty < t < \infty$. Let us find the function $a_1(t)$. Using the general expression (3.62) and transforming to the new variable $x = \tau - t$, we find

$$a_1(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{a(\tau)}{\tau - t} d\tau = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos \omega_0 \tau}{\tau - t} d\tau$$

$$= -\frac{1}{\pi} \cos \omega_0 t \int_{-\infty}^{\infty} \frac{\cos \omega_0 x}{x} dx + \frac{1}{\pi} \sin \omega_0 t \int_{-\infty}^{\infty} \frac{\sin \omega_0 x}{x} dx$$

It is known that the principal value of the integral

$$\int_{-\infty}^{\infty} \frac{\cos x}{x} dx = 0$$

and that

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} dx = \pi$$

Hence, the harmonic conjugate of the function $a(t) = \cos \omega_0 t$ is the function

$$a_1(t) = \sin \omega_0 t$$

which passes through zero when the original function passes through a maximum. In a similar manner, we may show that the function

$$a_1(t) = -\cos \omega_0 t$$

is the harmonic conjugate of the function $a(t) = \sin \omega_0 t$, $-\infty < t < \infty$.

Substituting $a(t) = \cos \omega_0 t$ and $a_1(t) = \sin \omega_0 t$ into expression 3.60, we obtain the well-known expression for the envelope of a harmonic wave:

$$A(t) = \sqrt{\cos^2 \omega_0 t + \sin^2 \omega_0 t} = 1$$

A similar result is obtained for $a(t) = \sin \omega_0 t$ and $a_1(t) = -\cos \omega_0 t$.

It could be seen that expression (3.60) defines the envelope in the form of a line tangent to the maxima of the original function, and in the case of a harmonic wave, it connects two adjacent maxima along the shortest path, i.e., along a straight line. Thus, expression (3.60) defines the "simplest" envelope. This property of expression (3.60) is also valid in the case of a compound signal, provided

that the envelope changes slowly, i.e., if the signal in question is narrow-band [see expressions (3.2) and (3.3)].

If the original signal is a sum of spectral components

$$a(t) = \sum_n (a_n \cos \omega_n t + b_n \sin \omega_n t) \tag{3.64}$$

its harmonic conjugate will then be

$$a_1(t) = \sum_n (a_n \sin \omega_n t - b_n \cos \omega_n t) \tag{3.65}$$

Series (3.65) is called the harmonic conjugate of series (3.64).

If the signal $a(t)$ is represented by the Fourier integral

$$a(t) = \frac{1}{\pi} \int_0^\infty [a(\omega) \cos \omega t + b(\omega) \sin \omega t]\, d\omega \tag{3.66}$$

the function $a_1(t)$ can be written in the form of an integral

$$a_1(t) = \frac{1}{\pi} \int_0^\infty [a(\omega) \sin \omega t - b(\omega) \cos \omega t]\, d\omega \tag{3.67}$$

which is the harmonic conjugate of integral (3.66).

The relation between the spectra of the functions $a(t)$ and $a_1(t)$ can be easily established.

Since the Hilbert transformation of a harmonic signal does not change its amplitude, it is clear that the spectral density $\mathbf{S}_1(\omega)$ of the harmonic conjugate $a_1(t)$ cannot differ in magnitude from the spectral density $\mathbf{S}(\omega)$ of the original function $a(t)$, while the phase characteristic of the spectrum $\mathbf{S}_1(\omega)$ does differ from that of the spectrum $\mathbf{S}(\omega)$. Comparison between expressions (3.67) and (3.66) shows that all the spectral components of the function $a_1(t)$ have a phase delay equal to 90° relative to the components of the function $a(t)$. Consequently, when $\omega > 0$ the spectral densities $\mathbf{S}_1(\omega)$ and $\mathbf{S}(\omega)$ are related by the relation

$$\mathbf{S}_1(\omega) = -i\mathbf{S}(\omega), \quad \omega > 0 \tag{3.68}$$

In the negative-frequency region we have

$$\mathbf{S}_1(\omega) = i\mathbf{S}(\omega), \quad \omega < 0 \tag{3.69}$$

In consequence of the change of the phase characteristic, the harmonic conjugate $a_1(t)$ can differ considerably in form from the original function $a(t)$.

Finally, it is easy to notice that if the original signal is written as

$$a(t) = A(t) \cos \psi(t) = A(t) \cos [\omega_0 t + \theta(t) + \theta_0] \tag{3.70}$$

where the envelope $A(t)$ is determined by relation (3.60), then its harmonic conjugate may be represented by a similar expression:

$$a_1(t) = A(t) \sin \psi(t) = A(t) \sin [\omega_0 t + \theta(t) + \theta_0] \tag{3.71}$$

This follows directly from expression (3.61) and Fig. 3.23.

After the harmonic conjugate $a_1(t)$ has been found, one can readily find from expressions (3.60) and (3.61) the envelope $A(t)$, phase $\psi(t)$ and, finally, the instantaneous frequency of a narrow-band signal:

$$\omega(t) = \frac{d\psi(t)}{dt} = \frac{d}{dt}\left[\arctan\frac{a_1(t)}{a(t)}\right] = \frac{a(t)\,a_1'(t) - a_1(t)\,a'(t)}{a^2(t) + a_1^2(t)} \tag{3.72}$$

By separating the invariable part ω_0 from the obtained frequency $\omega(t)$, we can write an expression for $\psi(t)$

$$\psi(t) = \omega_0 t + \theta(t) + \theta_0 \tag{3.73}$$

in which $\theta(t)$ does not contain terms linearly depending on time. Thus any arbitrariness in selecting the "centre frequency" ω_0 of the signal and, correspondingly, the function $\theta(t)$ is eliminated.

Let us explain the use of the Hilbert transform for defining the envelope, phase, and instantaneous frequency of a signal, taking the following example.

Let the signal be given in the form of a sum of two harmonic waves with close frequencies ω_1 and ω_2:

$$a(t) = A_1 \cos\omega_1 t + A_2 \cos\omega_2 t \tag{3.74}$$

and let it be required to represent $a(t)$ in the form

$$a(t) = A(t)\cos[\omega_0 t + \theta(t) + \theta_0] \tag{3.75}$$

The difference $|\Delta\omega| = |\omega_2 - \omega_1|$ is supposed to be so small as compared with $(\omega_2 + \omega_1)/2$ that the wave $a(t)$ might be regarded as a narrow-band signal.

What are $A(t)$, ω_0, and $\theta(t)$ in this case? It is difficult to find the structure of the envelope and phase of the resultant wave $a(t)$ directly from expression (3.74). Therefore, we shall use expressions (3.60) and (3.61).

The harmonic conjugate is

$$a_1(t) = A_1 \sin\omega_1 t + A_2 \sin\omega_2 t$$

Using formula (3.60), we find the envelope of the signal $a(t)$:

$$A(t) = \sqrt{(A_1\cos\omega_1 t + A_2\cos\omega_2 t)^2 + (A_1\sin\omega_1 t + A_2\sin\omega_2 t)^2}$$

$$= A_1\sqrt{1 + k^2 + 2k\cos\Delta\omega t} \tag{3.76}$$

where

$$k = A_2/A_1, \quad \Delta\omega = \omega_2 - \omega_1$$

in which case it is assumed for the sake of definiteness that $k < 1$ and $\Delta\omega > 0$.

The phase of the resultant oscillation is found from formula (3.61):

$$\psi(t) = \arctan\frac{a_1(t)}{a(t)} = \arctan\frac{\sin\omega_1 t + k\sin\omega_2 t}{\cos\omega_1 t + k\cos\omega_2 t} \tag{3.77}$$

Using formula (3.72) and performing simple algebraic and trigonometric transformations, we obtain the following expression for the instantaneous frequency:

$$\omega(t) = \omega_1 + \Delta\omega k \frac{k + \cos \Delta\omega t}{1 + k^2 + 2k \cos \Delta\omega t} = \omega_1 + \Delta\omega \eta(t) \tag{3.78}$$

where

$$\eta(t) = k \frac{k + \cos \Delta\omega t}{1 + k^2 + 2k \cos \Delta\omega t} \tag{3.79}$$

Since the d-c component of the function $\eta(t)$ is zero, the centre frequency ω_0 and the function $\theta(t)$ entering into expression (3.73) will be

$$\omega_0 = \omega_1 \tag{3.80}$$

$$\theta(t) = \Delta\omega \int_0^t \eta(t)\, dt \tag{3.81}$$

Thus, on the basis of (3.76), (3.78), (3.80), and (3.81), expression (3.75) is reduced to the form

$$a(t) = A_1 \sqrt{1 + k^2 + 2k \cos \Delta\omega t} \cos\left[\omega_1 t + \Delta\omega \int_0^t \eta(t)\, dt\right] \tag{3.82}$$

where $\eta(t)$ is defined by expression (3.79).

In this case any uncertainty and arbitrariness in selecting the envelope and phase of the resultant oscillation are eliminated.

The diagrams of the function $\eta(t)$, characterizing the change in frequency, for some values of k are shown in Fig. 3.24. For $k \to 1$ we have spikes described by delta functions. These spikes correspond to the derivatives of the stepwise phase changes (through 180°) occurring when the envelope of the beats goes to zero. When $k \ll 1$, i.e., when a weak oscillation $A_2 \cos \omega_2 t$ is superimposed on a strong one $A_1 \cos \omega_1 t$, expressions (3.76) through (3.79) are substantially simplified:

$$\left.\begin{aligned} A(t) &\approx A_1 (1 + k \cos \Delta\omega t), \quad \omega_0 = \omega_1 \\ \omega(t) &\approx \omega_1 + k\Delta\omega \cos \Delta\omega t \\ \psi(t) &\approx \omega_1 t + k \sin \Delta\omega t \end{aligned}\right\} \tag{3.83}$$

In this case the envelope, frequency, and phase of the resultant oscillation change harmonically with a frequency $|\Delta\omega| = |\omega_2 - \omega_1|$ relative to their mean values A_1, ω_1 and $\omega_1 t$, respectively.

Formulas (3.76) through (3.83) are of great practical importance, because radio engineers often have to deal with beats of two harmonic waves.

In conclusion, it should be noted that in some special cases expressions (3.60) through (3.69) are also used for wide-band signals where the concept of "envelope" can no longer be used in its usual sense.

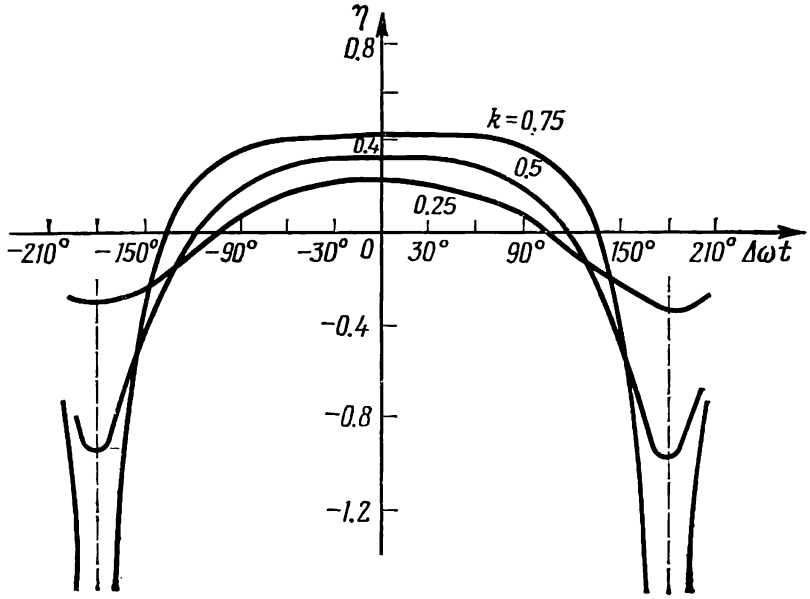


Fig. 3.24. Instantaneous frequency of an oscillation that is a sum of two harmonic oscillations

In this case one has to abandon the requirement that the envelope $A(t)$ should touch the curve $a(t)$ near the points where $a(t)$ has a maximum value.

3.10. ANALYTIC SIGNAL

In electrical engineering, a harmonic wave (current or voltage) is usually represented in the form

$$a(t) = A_0 \cos(\omega_0 t + \theta_0) = A_0 \operatorname{Re}[e^{i(\omega_0 t+\theta_0)}] = \operatorname{Re}[\mathbf{A}_0 e^{i\omega_0 t}] \quad (3.84)$$

or

$$a(t) = A_0 \sin(\omega_0 t + \theta_0) = A_0 \operatorname{Im}[e^{i(\omega_0 t+\theta_0)}] = \operatorname{Im}[\mathbf{A}_0 e^{i\omega_0 t}] \quad (3.84')$$

where $\mathbf{A}_0 = A_0 e^{i\theta_0}$ is the complex amplitude.

The symbols Re and Im are often omitted and the wave is simply expressed as

$$a(t) = A_0 e^{i[\omega_0 t+\theta_0]} = \mathbf{A}_0 e^{i\omega_0 t} \quad (3.85)$$

implying the real or the imaginary part of this expression.

In modern radio engineering the representation of oscillations in a complex form has been further advanced and extended to cover non-harmonic waves.

If a "physical" signal is determined as a real-valued function $a(t)$, the corresponding complex signal takes the form

$$z(t) = a(t) + ia_1(t) \tag{3.86}$$

where $a_1(t)$ is the Hilbert transform (harmonic conjugate) of the signal $a(t)$.

The complex function $z(t)$ defined in this way is called the complex or *analytic* signal corresponding to the physical signal $a(t)$. It should be noted that in expression (3.85) the imaginary part of the complex function is also the harmonic conjugate of the real part.

When $a(t)$ and $a_1(t)$ are represented in the form of expressions (3.70), (3.71), the analytic signal can be written as

$$z(t) = A(t)\, e^{i\psi(t)} = A(t)\, e^{i[\omega_0 t+\theta(t)+\theta_0]} = \mathbf{A}(t)\, e^{i\omega_0 t} \tag{3.87}$$

where

$$\mathbf{A}(t) = A(t)\, e^{i[\theta(t)+\theta_0]} \tag{3.88}$$

is the complex envelope of the narrow-band signal. The modulus of the complex envelope, equal to $A(t)$ [since $|e^{i[\theta(t)+\theta_0]}| = 1$ whatever the law governing the variation of $\theta(t)$] bears only the information of the amplitude modulation of the given wave, while the phase-defining factor $e^{i\theta(t)}$, only that of the angle modulation. But as a whole, the product $A(t)e^{i\theta(t)}$ contains all the information of the signal $a(t)$, except for the carrier frequency ω_0, which is supposed to be known.

This property of the complex envelope, allowing the frequency ω_0 to be disregarded when analysingnarrow-band signals, makes the concept of "analytic signal" very important.

Let us consider the main properties of the analytic signal and complex envelope.

1. *The product of the analytic signal $z_a(t)$ and the complex-conjugate signal $z_a^*(t)$ is equal to the square of the envelope of the original (physical) signal $a(t)$.*

Indeed

$$z_a(t)\, z_a^*(t) = [a(t) + ia_1(t)]\,[a(t) - ia_1(t)]$$
$$= a^2(t) + a_1^2(t) = A^2(t) \tag{3.89}$$

Thus, the modulus of the analytic signal $z_a(t)$ is equal to the envelope of the signal $z_a(t)$: $|z_a(t)| = A(t)$.

2. *The spectrum of the analytic signal contains positive frequencies only.*

From the expression

$$z_a(t) = a(t) + ia_1(t)$$

it follows that the spectral density $\mathbf{Z}(\omega)$ of the analytic signal $z_a(t)$ is defined by the sum

$$\mathbf{Z}_a(\omega) = \mathbf{S}_a(\omega) + i\mathbf{S}_{a_1}(\omega)$$

However, according to (3.68) and (3.69), when $\omega > 0$, $\mathbf{S}_{a_1}(\omega) = -i\mathbf{S}_a(\omega)$, while when $\omega < 0$, $\mathbf{S}_{a_1}(\omega) = i\mathbf{S}_a(\omega)$. Therefore

$$\mathbf{Z}_a(\omega) = \begin{cases} 2\mathbf{S}_a(\omega) & \text{when } \omega > 0 \\ 0 & \text{when } \omega < 0 \end{cases} \tag{3.90}$$

For example, if a narrow-band signal $a(t)$ has a spectral density $\mathbf{S}_a(\omega)$ whose modulus is shown in Fig. 3.25 by the dashed line, the analytic signal $z_a(t) = a(t) + ia_1(t)$ has a spectral density $\mathbf{Z}_a(\omega)$ whose modulus is shown in the same figure by the solid line.

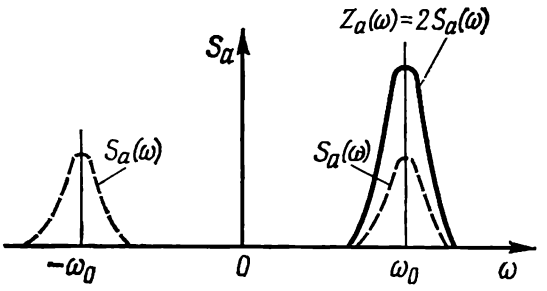


Fig. 3.25. Relation between the spectra of a physical and the analytic signal

The Fourier integral for the analytic signal $z_a(t)$ assumes the following form:

$$z_a(t) = \frac{1}{2\pi}\int_0^\infty \mathbf{Z}_a(\omega)e^{i\omega t}\,d\omega = \frac{1}{\pi}\int_0^\infty \mathbf{S}_a(\omega)\,e^{i\omega t}\,d\omega \tag{3.91}$$

where $\mathbf{S}_a(\omega)$ is the spectral density of the original (physical) signal $a(t)$*.

3. *The spectral density of the complex envelope* $\mathbf{A}(t)$ *coincides with the spectral density of the analytic signal* $z_a(t)$, *shifted to the left by an amount* ω_0.

Using the general formula (2.48), we may write

$$\mathbf{Z}_a(\omega) = \int_{-\infty}^{\infty} z_a(t)\,e^{-i\omega t}\,dt$$

Substituting $z_a(t) = \mathbf{A}(t)e^{i\omega_0 t}$ into this expression, we obtain

$$\mathbf{Z}_a(\omega) = \int_{-\infty}^{\infty} \mathbf{A}(t)\,e^{-i(\omega-\omega_0)t}\,dt = \mathbf{S}_A(\omega-\omega_0),\ \omega > 0 \tag{3.92}$$

This expression is a generalization of formula (2.58) to the case of a complex function of time, $\mathbf{A}(t)$, multiplied by $e^{i\omega_0 t}$ (instead of

* This expression makes clear the meaning of the term "analytic signal". In fact, the integral (3.91), when transformed to a complex t, converges in the upper half plane and is an analytic function at $\operatorname{Im} t > 0$, because when $\omega \to \infty$, the factor $e^{-i\omega \operatorname{Im} t}$ provides for convergence. On the other hand, in the case of a physical signal containing both positive and negative frequencies, the factor $e^{-i\omega \operatorname{Im} t}$ infinitely increases either when $\omega \to +\infty$ or when $\omega \to -\infty$.

$\cos \omega_0 t$ as in Sec. 2.7-3). Expression (3.9) given for the real envelope $A(t)$ (in purely amplitude modulation) is a particular case of the general expression (3.92).

Denoting $\omega - \omega_0 = \Omega$, let us rewrite expression (3.92) in a somewhat different form:

$$\mathbf{Z}_a(\omega_0 + \Omega) = \mathbf{S}_A(\Omega) = 2\mathbf{S}_a(\omega_0 + \Omega) \tag{3.93}$$

[see formula (3.90)].

The relation between the spectra $\mathbf{S}_A(\Omega)$ and $\mathbf{Z}_a(\omega_0 + \Omega)$ is illustrated in Fig. 3.26.

It should be emphasized that the spectrum $\mathbf{S}_A(\Omega)$ of the complex envelope $\mathbf{A}(t)$ is not necessarily symmetrical with respect to the zero frequency (see Fig. 3.26). If the spectrum $\mathbf{S}_a(\omega)$ of the physical signal $a(t)$ is asymmetrical with respect to $\omega = \omega_0$, as it may happen, for example, in combined amplitude-angle modulation (see Sec. 3.8), the function $\mathbf{Z}_a(\omega) = 2\mathbf{S}_a(\omega)$, $\omega > 0$, is also asymmetrical; after shifting $\mathbf{Z}_a(\omega)$ to the left by an amount ω_0, the spectrum $\mathbf{S}_A(\Omega)$ of the complex envelope will be asymmetrical relative to the frequency $\Omega = 0$. Since with such a shift the function $\mathbf{S}_A(\Omega)$ differs from zero in the region of frequencies $\Omega < 0$, the *complex function* $\mathbf{A}(t)$ *is not an analytic signal.* This is explained by the fact that the real and imaginary parts of $\mathbf{A}(t)$ do not form a Hilbert-transform pair.

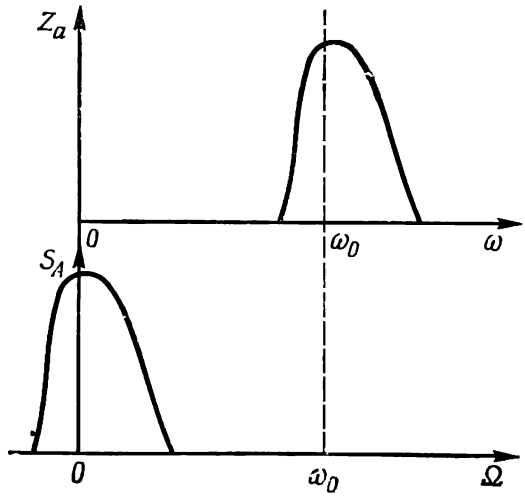


Fig. 3.26. Relation between the spectra of the analytic signal and the complex envelope of the original signal

4. *The autocorrelation function of the analytic signal*, defined by the expression

$$B_z(\tau) = \int_{-\infty}^{\infty} z_a^*(t)\, z_a(t+\tau)\, dt \tag{3.94}$$

is related to the auto-correlation function of the original physical signal (narrow-band), defined as

$$B_a(\tau) = \int_{-\infty}^{\infty} a(t)\, a(t+\tau)\, dt \tag{3.95}$$

by the relation

$$B_a(\tau) = \frac{1}{2} \operatorname{Re} B_z(\tau) \tag{3.96}$$

Indeed, substituting $z_a(t) = a(t) + ia_1(t)$ into (3.94), we obtain

$$B_z(\tau) = \int [a(t) + ia_1(t)]\,[a(t+\tau) - ia_1(t+\tau)]\,dt$$

$$= \int a(t)\,a(t+\tau)\,dt + \int_{-\infty}^{\infty} a_1(t)\,a_1(t+\tau)\,dt$$

$$+ i\left[\int_{-\infty}^{\infty} a_1(t)\,a(t+\tau)\,dt - \int_{-\infty}^{\infty} a(t)\,a_1(t+\tau)\,dt\right]$$

In Sec. 2.17 it was found that the autocorrelation function of a physical signal depends only on the modulus of its spectral density. Since the magnitudes of the spectra of the functions $a(t)$ and $a_1(t)$ are the same (see Sec. 3.9), the first two integrals are equal and add together, and the second two do not cancel out, hence relation (3.96). Now, substituting $z_a(t) = \mathbf{A}(t)\,e^{i\omega_0 t}$ and $z_a^*(t) = \mathbf{A}^*(t)\,e^{-i\omega_0 t}$ into (3.94), we have

$$B_z(\tau) = e^{-i\omega_0\tau} \int_{-\infty}^{\infty} \mathbf{A}(t)\,\mathbf{A}^*(t+\tau)\,dt \tag{3.97}$$

from which stems an important relation [see 3.(96)]

$$B_a(\tau) = \frac{1}{2} \operatorname{Re}\left[e^{-i\omega_0\tau} \int_{-\infty}^{\infty} \mathbf{A}(t)\,\mathbf{A}^*(t+\tau)\,dt\right] \tag{3.98}$$

The integral in (3.98) is the *autocorrelation function of the complex envelope* $\mathbf{A}(t)$. Therefore, expression (3.98) may be rewritten in the form

$$B_a(\tau) = \frac{1}{2} \operatorname{Re}\left[e^{-i\omega_0\tau} B_A(\tau)\right] \tag{3.99}$$

In particular, at $\tau = 0$, we get

$$B_a(0) = \frac{1}{2} \int_{-\infty}^{\infty} A^2(t)\,dt = \frac{1}{2} B_z(0) \tag{3.100}$$

From this expression it follows that, since $B_a(0) = E$, *the energy of the analytic signal is twice the energy of the original physical signal.*

It should be noted that the application of the concept of energy to a complex function is not entirely formal. In Chapter 12 we will see that in some signal processing systems one has to deal with

a combination of two functions of time that form a Hilbert-transform pair, i.e., with an analytic signal as a physical process.

The advantages of the analytic signal in the analysis of narrow-band processes will be clear from the following chapters.

3.11. AUTOCORRELATION FUNCTION OF A MODULATED WAVE

When finding the autocorrelation function of a modulated wave $a(t) = A(t) \cos \psi(t)$, we shall assume that the function $a(t)$ is absolutely integrable (possesses a finite energy), so that we may use the following definition (see Sec. 2.16):

$$B_a(\tau) = \int_{-\infty}^{\infty} a(t)\, a(t+\tau)\, dt \tag{3.101}$$

The computation of the integral for complex signals is fairly cumbersome. The problem can be substantially simplified by transforming from the wave $a(t)$ to the analytic signal $z_a(t) = \mathbf{A}(t) e^{i\omega_0 t}$. Using the relations derived in the preceding section, let us first consider purely amplitude modulation, when $a(t) = A(t) \cos \omega_0 t$, $\theta(t) = 0$ and, therefore, $\mathbf{A}(t) = \mathbf{A}^*(t) = A(t)$.

In this case, formula (3.98) assumes the following form

$$B_a(\tau) = \frac{1}{2} \operatorname{Re}\left[e^{-i\omega_0 \tau} \int_{-\infty}^{\infty} A(t)\, A(t+\tau)\, dt \right]$$

$$= \frac{1}{2} \cos \omega_0 \tau \int_{-\infty}^{\infty} A(t)\, A(t+\tau)\, dt \tag{3.102}$$

Designating, as in expression (3.99), the integral term by $B_A(\tau)$, we finally obtain

$$B_a(\tau) = B_A(\tau)\, (\cos \omega_0 \tau / 2) \tag{3.103}$$

The second factor $(\cos \omega_0 \tau/2)$ is the autocorrelation function of a harmonic wave of frequency ω_0 and unit amplitude.

Thus, the autocorrelation function of an amplitude-modulated radio signal is equal to the product of the autocorrelation functions of the envelope and the high-frequency carrier.

Shown as an example in Fig. 3.27*a* is a radio pulse with a square envelope, whose autocorrelation function is shown in Fig. 3.27*b*. It should be noted that this function is independent of the epoch angle of the carrier, while its envelope coincides with the autocorrelation function of a square video pulse (see Sec. 2.16, Fig. 2.38*d*).

To illustrate the application of expression (3.99) to the case of a combined amplitude-frequency modulation, let us find the autocorrelation function of the pulse shown in Fig. 3.19*a*.

When using the symbols given in formula (3.37) and in Fig. 3.19, the analytic signal is written in the form

$$z_a(t) = A_0 e^{i\beta t^2/2} e^{i\omega_0 t}, \quad -T_s/2 \leqslant t \leqslant T_s/2 \tag{3.104}$$

Using formulas (3.94) and (3.99), we get

$$B_a(\tau) = \frac{A_0^2}{2} \operatorname{Re} \int_{-T_s/2}^{T_s/2-\tau} e^{i[\omega_0 t + \beta t^2/2]} e^{-i[\omega_0 (t+\tau) + \beta (t+\tau)^2/2]} dt \tag{3.105}$$

The limits of integration are taken with consideration for the condition of simultaneous existence of the functions $a(t)$ and

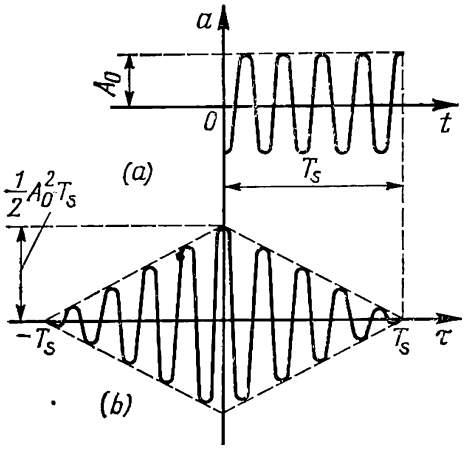


Fig. 3.27. (*a*) Pulse with a radio-frequency carrier and (*b*) its autocorrelation function

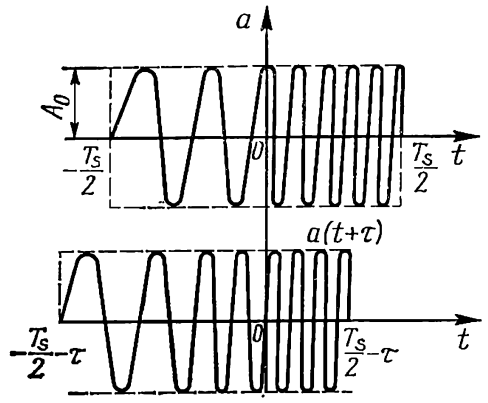


Fig. 3.28. Plotting the autocorrelation function of a linear FM pulse

$a(t+\tau)$ (Fig. 3.28).

Performing simple transformations, we may reduce expression (3.105) to the form

$$B_a(\tau) = \begin{cases} \dfrac{A_0^2 \sin\left(\dfrac{\beta T_s}{2}\tau - \dfrac{\beta\tau^2}{2}\right)\cos\omega_0\tau}{\beta\tau} & \text{for } |\tau| \leqslant T_s/2 \\ 0 & \text{for } |\tau| > T_s/2 \end{cases} \tag{3.106}$$

Using the parameter m introduced in Sec. 3.7 [see formula (3.38)] and taking into account that $\beta T_s^2 = 2\omega_d T_s = 2\pi m$, we may reduce expression (3.106) to a more general form

$$B_a(\tau) = \frac{1}{2} A_0^2 T_s \frac{\sin\left[\pi m \dfrac{\tau}{T_s}\left(1 - \dfrac{\tau}{T_s}\right)\right]}{\dfrac{\pi m \tau}{T_s}} \cos\omega_0\tau \tag{3.106'}$$

The factor $A_0^2 T_s/2 = B_a(0) = E$ is equal to the total energy of the radio pulse, as in the case of a pulse with a constant carrier frequency (see Fig. 3.27*b*).

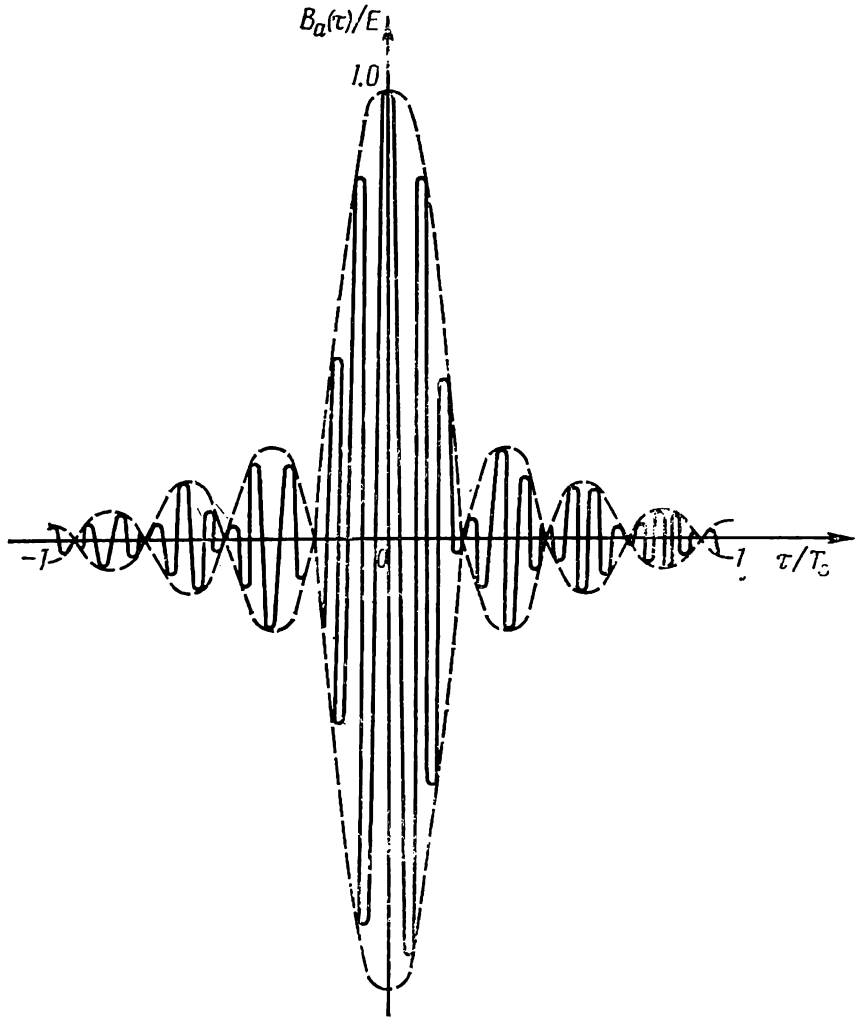


Fig. 3.29. Autocorrelation function of a linear FM pulse

Thus,

$$\frac{B_a(\tau)}{B_a(0)} = \frac{B_a(\tau)}{E} = \frac{\sin\left[\frac{\pi m\tau}{T_s}\left(1-\frac{\tau}{T_s}\right)\right]}{\frac{\pi m\tau}{T_s}} \cos\omega_0\tau \qquad (3.107)$$

The graph of this function for $m = 100$ is given in Fig. 3.29 under assumption that $\omega_0 T_s$ is very large (in Fig. 3.29 an arbitrary scale is used). The envelope of the autocorrelation function forms a fairly sharp peak (for $m \gg 1$), while the variation frequency of the curve

inscribed in the envelope is constant and equal to the centre frequency ω_0 of the original radio pulse.

The above signal and its autocorrelation function are of great interest in modern radio engineering.

3.12. SAMPLING OF A NARROW-BAND SIGNAL

Let us take a signal

$$a(t) = A(t) \cos \psi(t) = A(t) \cos [\omega_0 t + \theta(t)] \qquad (3.108)$$

whose spectrum is contained within a narrow frequency band from ω_1 to ω_2, so that the modulus of the spectral density, $S_a(\omega)$, has the

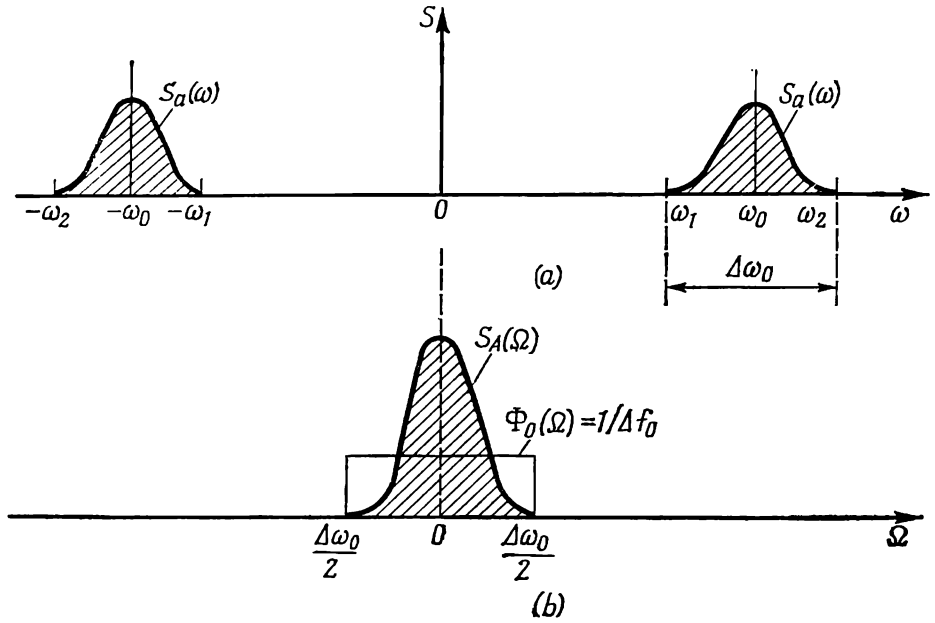


Fig. 3.30. Spectrum of (*a*) a narrow-band signal and (*b*) the complex envelope of this signal

form shown in Fig. 3.30*a*, the spectrum in the frequency band $\Delta\omega_0$ being not necessarily symmetrical with respect to the centre frequency $\omega_0 = (\omega_1 + \omega_2)/2$. For the signal to be narrow-band the following condition must be satisfied:

$$\Delta\omega_0/\omega_0 = \Delta f_0/f_0 \ll 1$$

where $\Delta f_0 = \Delta\omega_0/2\pi = f_2 - f_1$ is the frequency band, Hz.

The function $A(t)$ is supposed to be the "simplest" envelope, i.e., that $A(t)$ and $\psi(t)$ satisfy relations (3.60) and (3.61).

If the sampling of such a signal is based on series (2.120), the spacing between the samples must not exceed $1/(2f_2)$, where f_2 is the highest frequency in the signal spectrum.

The inexpediency of such an approach is evident, since the information about the signal is contained not in the frequency f_2 (or f_1) but in the envelope $A(t)$ or phase $\theta(t)$, which slowly change in time at relatively low modulation frequencies.

Therefore, it is desirable to transform expression (2.120) so that the intervals between the samples are defined by the actual width of the spectrum, i.e., by the quantity Δf_0 and not by the upper frequency f_2.

For this purpose, let us transform to the analytic signal corresponding to the given function $a(t)$:

$$z_a(t) = A(t)\, e^{i\psi(t)} = A(t)\, e^{i\theta(t)} e^{i\omega_0} = \mathbf{A}(t)\, e^{i\omega_0 t} \tag{3.109}$$

where the complex envelope $\mathbf{A}(t) = A(t)e^{i\theta(t)}$ is a low-frequency function whose spectrum $\mathbf{S}_A(\Omega)$ adjoins the zero frequency (Fig. 3.30*b*). Let the complex function $\mathbf{A}(t) = A(t)e^{i\theta(t)}$ be expanded in the orthogonal set

$$\mathbf{A}(t) = \sum_{n=-\infty}^{\infty} c_n \varphi_n(t) \tag{3.110}$$

where the basic function $\varphi_n(t)$ is defined by expression (2.121).

Substituting this series into (3.109), we obtain

$$z_a(t) = \left[\sum_{n=-\infty}^{\infty} c_n \varphi_n(t)\right] e^{-i\omega_0 t} \tag{3.111}$$

Now, let us define the original oscillation $a(t)$ as the real part of the function $z_a(t)$:

$$a(t) = \mathrm{Re}\left\{\left[\sum_{n=-\infty}^{\infty} c_n \varphi_n(t)\right] e^{i\omega_0 t}\right\} \tag{3.112}$$

As is seen, the problem of sampling a radio-frequency oscillation is reduced to sampling the complex envelope $\mathbf{A}(t)$. When determining the largest admissible interval between the samples in expansion (3.110), it is necessary to proceed from the highest frequency in the spectrum of the function $\mathbf{A}(t)$. From the definition of ω_0 as the centre frequency in the band $\Delta\omega_0$ it is clear that this frequency, as read from $\Omega = 0$, is equal to $\Delta\omega_0/2$, or $\Delta f_0/2$ (in Hz). Therefore, the spacing between the samples must not exceed

$$\Delta t = 1/2\ (\Delta f_0/2) = 1/\Delta f_0 \tag{3.113}$$

and the function $\varphi_n(t)$ must have the form

$$\varphi_n(t) = \frac{\sin(\Delta\omega_0/2)(t - n\Delta t)}{(\Delta\omega_0/2)(t - n\Delta t)} = \frac{\sin \pi\Delta f_0 (t - n\Delta t)}{\pi\Delta f_0 (t - n\Delta t)} \tag{3.114}$$

The function $\varphi_n(t)$ differs from the similar function used in Sec. 2.14 only in that ω_m here is replaced by $\Delta\omega_0/2$. Hence, the

spectral density $\Phi_0(\Omega)$ of the function $\varphi_0(t)$ is equal to $2\pi/\Delta\omega_0 = 1/\Delta f_0$ in the frequency band $|\Omega| \leqslant \Delta\omega_0/2$ (Fig. 3.30), while that of the function $\varphi_n(t)$ is given by

$$\Phi_n(\Omega) = \begin{cases} \frac{\pi}{\Delta\omega_0/2} e^{-in\Delta t\Omega} = \frac{1}{\Delta f_0} e^{-in\Delta t\Omega} & \text{for } |\Omega| \leqslant \frac{\Delta\omega_0}{2} \\ 0 & \text{for } |\Omega| > \frac{\Delta\omega_0}{2} \end{cases} \tag{3.115}$$

By analogy with expression (2.123), the square of the norm of the function $\varphi_n(t)$ is

$$\| \varphi_n \|^2 = \pi/0.5\, \Delta\omega_0 = 1/\Delta f_0 \tag{3.116}$$

Further, using formula (2.9) and taking into account (3.116), we have

$$c_n = \frac{1}{\| \varphi_n \|^2} \int_{-\infty}^{\infty} \mathbf{A}(t)\, \varphi_n\, dt = \Delta f_0 \int_{-\infty}^{\infty} \mathbf{A}(t)\, \varphi_n(t)\, dt \tag{3.117}$$

Using formula (2.63) and replacing ω by Ω^*, we get

$$c_n = \Delta f_0 \frac{1}{2\pi} \int_{-\Delta\omega_0/2}^{\Delta\omega_0/2} \mathbf{S}_A(\Omega)\, \Phi_n(-\Omega)\, d\Omega$$

$$= \Delta f_0 \frac{1}{2\pi} \int_{-\Delta\omega_0/2}^{\Delta\omega_0/2} \mathbf{S}_A(\Omega) \frac{1}{\Delta f_0} e^{in\Delta t\Omega}\, d\Omega$$

$$= \mathbf{A}(n\Delta t) = A(n\Delta t)\, e^{i\theta(n\Delta t)} \tag{3.118}$$

In expression (3.118), $\mathbf{S}_A$ is the spectrum of the complex envelope $\mathbf{A}(t)$ and $A(n\Delta t)$ is its magnitude at the sampling point $t = n\Delta t$. Thus, the coefficients of series (3.110) are samples of the function $\mathbf{A}(t)$ taken at intervals $\Delta t = 1/\Delta f_0$.

Substituting (3.118) into (3.111), we obtain

$$z_a(t) = \sum_{n=-\infty}^{\infty} A(n\Delta t)\, \varphi_n(t)\, e^{i[\omega_0 t + \theta(n\Delta t)}$$

and, using formula (3.112), define

$$a(t) = \sum_{n=-\infty}^{\infty} A(n\Delta t)\, \varphi_n(t) \cos[\omega_0 t + \theta(n\Delta t)]$$

$$= \sum_{n=-\infty}^{\infty} A(n\Delta t) \frac{\sin \pi\Delta f_0 (t - n\Delta t)}{\pi\Delta f_0 (t - n\Delta t)} \cos[\omega_0 t + \theta(n\Delta t)] \tag{3.119}$$

* Because we consider here the spectrum of the envelope.

At a given signal duration T_s the number of sampling points $T_s/\Delta t = T_s \Delta f_0$, two parameters — $A(n\Delta t)$ and $\theta(n\Delta t)$ — being necessarily specified at each point.

It should be noted that in the case of an asymmetrical spectrum (in the frequency band $\Delta\omega_0$) the frequency $\omega_0 = (\omega_1 + \omega_2)/2$ introduced in this section may not coincide with the "centre" frequency in expression (3.73). In other words, the phase $\theta(t)$ may contain a term that linearly depends on time.

Let us illustrate expression (3.119), taking an amplitude- or a frequency-modulated wave as an example.

In the case of amplitude modulation we proceed from the oscillation $a(t) = A(t) \cos \omega_0 t$, where $A(t)$ is a real function with a spectrum $S_A(\omega)$ limited by the highest frequency $\Omega_m = 2\pi F_m$. In this case the width of the spectrum of the modulated wave $a(t)$ is equal to $\Delta f_{am} = 2F_m$, the spectral density $S_a(\omega)$ being symmetrical relative to ω_0 within the frequency band. In accordance with formula (3.113), the interval between the samples must not exceed $\Delta t = 1/\Delta f_{am} = 1/2\ F_m$, i.e., it must be exactly the same as when sampling the original message (the modulating voltage).

Since the phase of the high-frequency carrier in purely amplitude modulation is constant, it need not be transmitted. Thus we arrive at the following conclusion: *an amplitude-modulated wave is fairly explicitly defined by the magnitudes of its amplitude sampled at intervals* $1/2F_m$, where F_m is the *upper frequency in the spectrum of the modulating function* (i.e., in the spectrum of the message being transmitted).

In other words, in the case of purely amplitude modulation, the number of degrees of freedom of the modulated wave is the same as that of the modulating function.

Now let us consider a frequency-modulated wave

$$a(t) = A_0 \cos [\omega_0 t + \theta(t)]$$

where the instantaneous frequency $\omega(t) = \omega_0 + d\theta/dt$ is modulated by the same message as in the preceding case, the maximum frequency deviation f_d being large compared with F_m so that the band width Δf_{fm} of the modulated wave may be considered to be equal to $2f_d$ [the case of "wide-band" frequency modulation (3.34)]. The interval between the samples must be taken equal to $\Delta t \leqslant 1/\Delta f_{fm} = 1/2\ f_d$. Since the wave amplitude in frequency modulation is constant, it need not be transmitted. Therefore, for unambiguous representation of a frequency-modulated wave, it is sufficient to specify the phase $\theta(n\Delta t)$ of this wave at the sampling points spaced at time intervals $\Delta t \leqslant 1/2\ f_d$. The duration T_s of the message being the same, the number of phase samples in frequency modulation is equal to $\Delta f_{fm} T_s = 2f_d T_s$, while that of envelope samples in amplitude modulation is equal to $\Delta f_{am} T_s = 2F_m T_s$. From this it is clear

that the number of degrees of freedom of a frequency-modulated signal is $f_d/F_m = m$ times greater than that of an amplitude-modulated signal, the transmitted message (amount of information) being the same. This is the result of the widening of the spectrum in frequency modulation. At the receiving end of a communication channel, after frequency detection of the modulated wave, a voltage is separated, whose spectrum and number of degrees of freedom are the same as those of the original message.

This example shows that with one and the same spectrum width, the information capacity of a radio signal differs with the type of modulation.

In the case of a combined modulation (amplitude and angle), two samples — amplitude and phase — must be taken at each sampling point.

Chapter 4

MAIN CHARACTERISTICS OF RANDOM SIGNALS

4.1. RANDOM PROCESSES. GENERAL

Information transmitted through a communication channel or derived through measurements is contained in a signal.

Before receiving the message (or before taking measurements) the signal should be regarded as a random process, that is, a set (ensemble) of functions of time, which show some statistical regularity. One of the functions which has become completely known, after receiving the message is called a sample function of realization of the random process. This realization is already not a random, but a deterministic function of time.

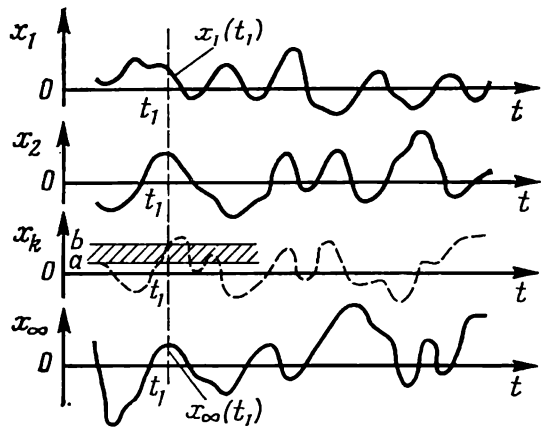


Fig. 4.1. Set of functions forming a random process

The one-dimensional probability distribution associated with a given random process is an important, but not exhaustive, characteristic of this process. Figure 4.1 shows a set of functions $x_1(t)$, $x_2(t)$ . . . forming a random process $X(t)$. The values that the individual functions can assume at a time $t = t_1$ form a set of random variables $x_1(t_1)$, $x_2(t_1)$, . . .

The probability that a random variable $x_k(t_1)$ takes a value in any given interval (a, b) is defined by the expression

$$P_{t_1}(a < x \leqslant b) = \int_a^b p(x; t_1)\, dx \qquad (4.1)$$

The function $p(x; t_1)$ is the differential distribution function of the random variable $x(t_1)$; $p(x; t_1)$ is called univariate (one-dimensional) probability density, and P_{t_1} is the integral probability.

The function $p(x; t_1)$ is valid for random x of continuous type, which can assume any value within a certain interval. No matter what the character of the function $p(x; t_1)$ is, the following equality

must hold:

$$\int_{x_{\min}}^{x_{\max}} p(x; t_1)\, dx = 1 \tag{4.2}$$

where $x_{\min}$ and $x_{\max}$ are the limits of the possible values of $x(t_1)$.

If x is a discrete random variable that can take any one of a finite set of discrete values, expression (4.2) must be replaced by the sum

$$\sum_i P_i = 1 \tag{4.2'}$$

where P_i is the probability associated with the value x_i.

Setting the univariate probability density $p(x; t_1)$ allows one to carry out the statistical averaging of any function $f(x)$, as well as of the variable x itself. Under statistical averaging is meant the averaging of x over a set of values (ensemble) in some "section" of the process, i.e., at a fixed instant of time.

For practical applications the following parameters of a random process are important:

1. The mean value (mathematical expectation, the first moment)

$$\langle x(t_1)\rangle = \int_{-\infty}^{\infty} xp(x; t_1)\, dx \tag{4.3}$$

2. The mean-square (the second moment)

$$\langle x^2(t_1)\rangle = \int_{-\infty}^{\infty} x^2 p(x; t_1)\, dx \tag{4.4}$$

3. The mean-square deviation (variance)

$$\sigma_x^2(t_1) = \langle (x - \langle x\rangle)^2\rangle = \langle x^2\rangle - (\langle x\rangle)^2 \tag{4.5}$$

In expressions (4.3) through (4.5) the pointed brackets $\langle\ \rangle$ imply the operation of ensemble averaging.

The univariate probability density is insufficient for complete description of a random process, since it gives a probabilistic idea of a random process $X(t)$ only at some fixed instants of time. The bivariate (two-dimensional) probability density $p(x_1, x_2; t_1, t_2)$*, making it possible to take into account the relation between the values x_1 and x_2 taken by a random function at arbitrarily selected instants of time, t_1 and t_2, is a more comprehensive characteristic of the process.

* Here and elsewhere the same letter p stands for the probability density of various random functions. In some sections, should it prove necessary to avoid confusion, we shall use subscripts denoting the parameter a given distribution is associated with. For example, when considering a random process $x(t) = A(t)\cos\theta(t)$, we shall use definitions $p_x(x)$, $p_A(A)$, and $p_\theta(\theta)$.

The n-variate (n-dimensional) probability density at sufficiently large n is an exhaustive probability characteristic of a random process. However, a great number of problems associated with the description of random signals can be solved using the bivariate probability density.

In particular, the use of the bivariate probability density $p(x_1, x_2; t_1, t_2)$ makes it possible to determine an important characteristic of a random process — the *autocorrelation function* (the second mixed moment)

$$B_x(t_1, t_2) = \langle x(t_1)\, x(t_2) \rangle \tag{4.6}$$

According to this definition, the autocorrelation function of a random process $X(t)$ is the statistically averaged product of the values of the random function $X(t)$ at the instants t_1 and t_2.

For each realization of a random process, the product $x(t_1)\, x(t_2)$ is a certain number. The totality of realizations forms a set of random numbers whose distribution is characterized by the bivariate probability density $p(x_1, x_2; t_1, t_2)$. For a given function $p(x_1, x_2; t_1, t_2)$, the operation of ensemble averaging, denoted in expression (4.6) by pointed brackets, is carried out by the formula

$$B_x(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1; x_2; t_1, t_2)\, dx_1\, dx_2 \tag{4.7}$$

When $t_2 = t_1$, the bivariate $x_1 x_2$ degenerates to a univariate $x_1^2 = x_2^2$. Therefore, in accordance with expression (4.4), we may write

$$B_x(t_1, t_1) = \int_{-\infty}^{\infty} x_1^2 p(x_1; t_1)\, dx_1 = \langle x_1^2 \rangle \tag{4.8}$$

Thus, with a zero interval between the instants t_1 and t_2, the autocorrelation function defines the mean-square value of the random process at the instant $t = t_1$.

The study of random processes and their action on radio circuits is substantially simplified if these processes are *stationary*.

A random process is called strictly stationary if its probability density $p(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$ of an arbitrary order n depends only on the intervals $t_2 - t_1$, $t_3 - t_1$, $\ldots$, $t_n - t_1$ and is independent of the position of these intervals within the range of the argument t.

In radio engineering applications of the theory of random processes the condition of stationarity is usually restricted to the requirement that only the univariate and bivariate probability densities should not depend on time (a wide-sense stationary random process). Satisfying this condition allows one to assume that the mean value

(the first moment), the mean square, and the variance of the given random process do not depend on time, while the autocorrelation function depends not on the instants t_1 and t_2 themselves, but on the interval $\tau = t_2 - t_1$ between them.

The wide-sense stationarity of a process may be treated as stationarity within the framework of the correlation theory (for the moments of order not exceeding two).

Thus, for a wide-sense stationary random process, expressions (4.3) through (4.5) and (4.7) may be written without denoting the fixed instants of time:

$$\langle x \rangle = \int_{-\infty}^{\infty} xp(x)\,dx \tag{4.9}$$

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 p(x)\,dx \tag{4.10}$$

$$\sigma_x^2 = \langle x^2 \rangle - (\langle x \rangle)^2 \tag{4.11}$$

$$B_x(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2, \tau)\,dx_1\,dx_2 \tag{4.12}$$

Further simplification of the analysis of random processes can be achieved by using the notion of *ergodicity*. A stationary random process is called ergodic if the ensemble averaging of any of its probability characteristics is equivalent to the time averaging of a single, theoretically infinitely long realization of this process.

The condition of ergodicity of a random process also includes the condition of its stationarity. In accordance with the definition of ergodic process, relations (4.9) through (4.12) are equivalent to the following expressions, in which the time averaging operation is denoted by the vinculum:

$$\overline{x(t)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\,dt \tag{4.13}$$

$$\overline{x^2(t)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)\,dt \tag{4.14}$$

$$\sigma_x^2 = \overline{x^2(t)} - [\overline{x(t)}]^2 \tag{4.15}$$

$$B_x(\tau) = \overline{x(t)\,x(t+\tau)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\,x(t+\tau)\,dt \tag{4.16}$$

If $x(t)$ is an electric signal (current, voltage), $\overline{x(t)}$ is then the d-c component of the random signal, $\overline{x^2(t)}$ is the average power, and

σ_x^2 is the average power of the fluctuation of the signal [with respect to the d-c component $\overline{x(t)}$].

Expression (4.16) externally coincides with the definition (2.134) of the autocorrelation function of a periodic deterministic signal. The even parity of the function $B_x(\tau)$ with respect to the time shift τ stems directly from (4.16).

It is also obvious that

$$B_x(0) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)\,dt = \overline{x^2(t)} = (\bar{x})^2 + \sigma_x^2 \qquad (4.17)$$

i.e., the value of the autocorrelation function at $\tau = 0$ is equal to the total average power of the random signal.

When analyzing a random processes, its fluctuation component is often of primary importance. In this case, use is made of the *autocovariance* function

$$\rho_x(\tau) = \overline{[x(t) - \overline{x(t)}]\,[x(t+\tau) - \overline{x(t)}]} = B_x(\tau) - (\bar{x})^2 \qquad (4.18)$$

Finally, the following *normalized* autocorrelation function is introduced:

$$R_x(\tau) = \frac{\rho_x(\tau)}{\sigma_x^2} = \frac{B_x(\tau) - (\bar{x})^2}{\sigma_x^2} \qquad (4.19)$$

The function $R_x(\tau)$ characterizes the correlation between the values $[x(t) - \bar{x}]$ at time τ apart. The slower and smoother the change of $x(t)$ in time, the longer the interval τ within which a statistical relation between the instantaneous values of the random function is observed.

In experimental study of random processes, use is made of the time correlation characteristics of the process, (4.13) through (4.19), because, as a rule, the experimenter can observe only a single realization of a given signal and not a set of its realizations. Naturally, integration is performed over a finite interval T, and the more stringent the requirements for the accuracy of measurements, the longer must be this interval.

4.2. TYPES OF RANDOM PROCESSES. EXAMPLES

To illustrate the general definitions given in the preceding section, let us consider several typical random processes.

Along with the symbol $X(t)$, we shall use $x(t)$ to denote a random function of time. As before, $x_k(t)$ stands for the kth realization of the random function $x(t)$.

4.2-1. A Signal in the Form of a Random Level D-C Voltage

Let the level A of the signal assume any value in the interval from $-A_{\max}$ to $A_{\max}$ (Fig. 4.2) with uniform probability.

It is obvious that this process is stationary. The univariate probability density can easily be obtained from expression (4.2):

$$p(x) = 1/2\ A_{\max},\quad -A_{\max} < x < A_{\max} \tag{4.20}$$

Substitution of (4.20) into expressions (4.3) through (4.5) gives

$$\langle x\rangle = \frac{1}{2A_{\max}} \int_{-A_{\max}}^{A_{\max}} x\,dx = 0$$

$$\langle x^2\rangle = \frac{1}{2A_{\max}} \int_{-A_{\max}}^{A_{\max}} x^2\,dx = \frac{1}{3} A_{\max}^2$$

$$\sigma_x^2 = \langle x^2\rangle - (\langle x\rangle)^2 = \frac{1}{3} A_{\max}^2$$

When determining the autocorrelation function by formula (4.6), one should take into account that for any realization, regardless of the time shift τ, the equality $x_k(t_1) = x_k(t_2)$ holds. Therefore, $x_k(t_1)\,x_k(t_2) = x_k^2$, and the ensemble averaging of this product can be carried out directly by formula (4.8) without using the bivariate probability density. Thus.

$$B_x(\tau) = \langle x^2\rangle = \frac{1}{3} A_{\max}^2 = \sigma_x^2$$

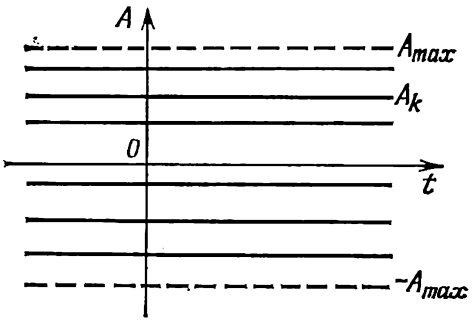


Fig. 4.2. Set of d-c voltages of random level

The process in question is *not ergodic*. This is clear from the fact that

$$\overline{x_k(t)} = x_k \neq \langle x\rangle,\quad \overline{x_k^2(t)} = x_k^2 \neq \langle x^2\rangle$$

$$B_x(\tau) = x_k^2 \neq \frac{1}{3} A_{\max}^2$$

i.e., the first two moments and the autocorrelation function determined by time averaging (along the realizations) do not coincide with the results of ensemble averaging.

4.2-2. Harmonic Oscillation with a Random Amplitude

Let a signal be defined by the expression

$$x(t) = A\cos(\omega_0 t + \theta_0) = A\cos\psi(t) \tag{4.21}$$

where the frequency ω_0 and epoch angle θ_0 are deterministic and constant, while the amplitude A is a random variable, equiprobable in the interval from 0 to $A_{\max}$ (Fig. 4.3).

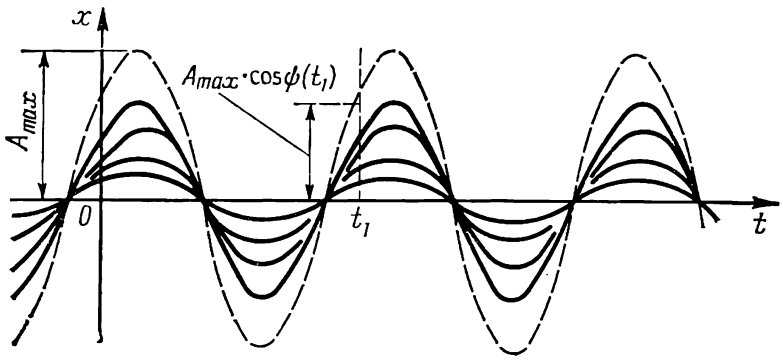


Fig. 4.3. Set of harmonic oscillations of random amplitude

Let us find the univariate probability density $p(x;\ t_1)$ for a fixed instant of time, t_1. The instantaneous value $x(t_1)$ can be anything within the interval from 0 to $A_{\max}\cos\psi(t_1)$, it being assumed that $\cos\psi(t_1) > 0$. Consequently,

$$p(x;\ t_1) = 1/A_{\max}\cos\psi(t_1),\quad 0 < x < A_{\max}\cos\psi(t_1)$$

and the mathematical expectation

$$\langle x(t_1)\rangle = \frac{1}{A_{\max}\cos\psi(t_1)}\int_0^{A_{\max}\cos\psi(t_1)} x\,dx = \frac{1}{2}A_{\max}\cos\psi(t_1)$$

Then

$$\langle x^2(t_1)\rangle = \frac{1}{A_{\max}\cos\psi(t_1)}\int_0^{A_{\max}\cos\psi(t_1)} x^2\,dx = \frac{1}{3}A_{\max}^2\cos^2\psi(t_1)$$

Finally, the variance is

$$\sigma_x^2(t_1) = \langle x^2(t_1)\rangle - [\langle x(t_1)\rangle]^2$$

$$= \frac{1}{3}A_{\max}^2\cos^2\psi(t_1) - \frac{1}{4}A_{\max}^2\cos^2\psi(t_1) = \frac{1}{12}A_{\max}^2\cos^2\psi(t_1)$$

The process under consideration is *not stationary* and *not ergodic*.

4.2-3. Harmonic Oscillation with a Random Phase

Let the amplitude A_0 and frequency ω_0 of a harmonic signal be explicitly known, and let the epoch angle θ be a random variable that can assume any value within the range from $-\pi$ to π with

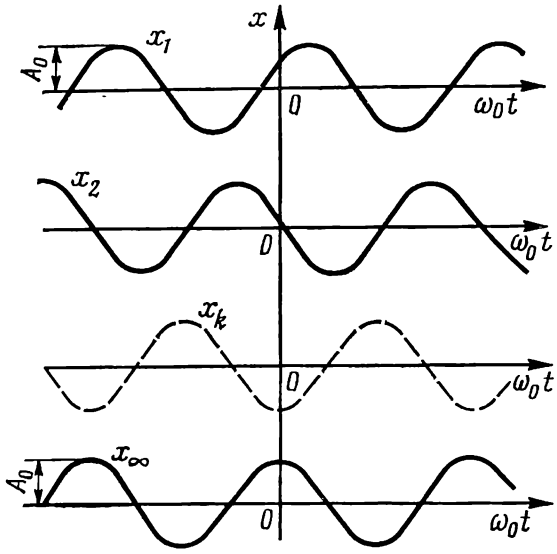


Fig. 4.4. Set of harmonic oscillations of random phase

uniform probability. This means that the probability density of the epoch angle is

$$p_\theta(\theta) = 1/2\pi, \quad -\pi < \theta < \pi \tag{4.22}$$

One of the realizations of the random process $x(t)$ formed by a set of harmonic waves with random phases (Fig. 4.4) may be defined by the expression

$$x_k(t) = A_0 \cos(\omega_0 t + \theta_k) = A_0 \cos \psi_k(t) \tag{4.23}$$

The phase angle $\psi(t) = \omega_0 t + \theta$ of the wave is a random variable, equiprobable within the range from $\omega_0 t - \pi$ to $\omega_0 t + \pi$, hence,

$$p_\psi(\psi) = 1/2\pi, \quad \omega_0 t - \pi < \psi < \omega_0 t + \pi \tag{4.24}$$

Let us find the univariate probability density $p_x(x)$ of the random process $X(t)$. Let us take an interval x, $x + dx$ (Fig. 4.5) and determine the probability that the instantaneous value of the signal falls in this interval, when being measured during a time interval from t_1 to $t_1 + dt$. This probability may be written in the form $p_x(x)\,dx$, where $p_x(x)$ is the sought probability density. It is clear that the probability $p_x(x)\,dx$ coincides with the probability that the random

phase ψ of the waves falls in one of the two hatched phase intervals in Fig. 4.5. This latter probability is equal to $2p_\psi(\psi)\,d\psi$. Consequently,

$$p_x(x)\,dx = 2p_\psi(\psi)\,d\psi = (2/2\pi)\,d\psi,$$

whence the sought function*

$$p_x(x) = \frac{1}{\pi}\frac{1}{\left|\frac{dx}{d\psi}\right|}, \quad -A_0 < x < A_0$$

But

$$\left|\frac{dx}{d\psi}\right| = A_0\,|\sin\psi| = A_0\sqrt{1-\cos^2\psi} = \sqrt{A_0^2 - x^2}$$

Thus, finally, we have

$$p_x(x) = 1/\pi\sqrt{A_0^2 - x^2}, \quad -A_0 \leqslant x \leqslant A_0 \tag{4.25}$$

The graph of this function is shown in Fig. 4.6.

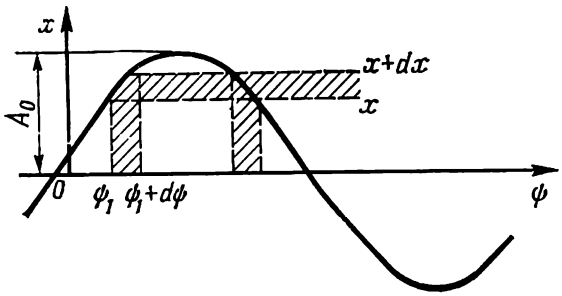


Fig. 4.5. Determining the probability density of a harmonic oscillation of random phase

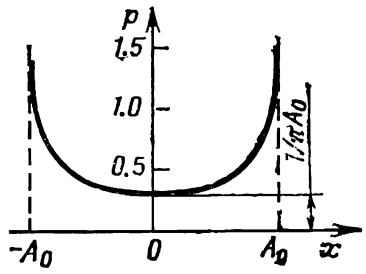


Fig. 4.6. Probability density of harmonic oscillation of random phase

It is essential that the univariate probability density $p_x(x)$ is independent of the choice of the instant t, while the ensemble average (see (2.271.7) in [10]):

$$\langle x(t)\rangle = \int_{-A_0}^{A_0} x p_x(x)\,dx = \frac{1}{\pi}\int_{-A_0}^{A_0}\frac{x}{\sqrt{A_0^2 - x^2}}\,dx = 0 \tag{4.26}$$

coincides with the time average

$$\overline{x(t)} = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x(t)\,dt = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} A_0\cos(\omega_0 t + \theta)\,dt = 0$$

(This holds true for any realization of the random process in question.)

* The absolute value of the derivative is taken on the grounds that the probability density is a non-negative function.

In this case, the autocorrelation function can be obtained by finding the ensemble average of the product $x(t_1)\,x(t_2)$ without using the bivariate probability density [see general expression (4.7)]. In fact, by substituting

$$x(t_1)\,x(t_2) = A_0^2 \cos(\omega_0 t_1 + \theta)\cos(\omega_0 t_2 + \theta)$$
$$= \frac{1}{2} A_0^2 \{\cos\omega_0 (t_2 - t_1) + \cos[\omega_0 (t_1 + t_2) + 2\theta]\}$$

into (4.6) and taking into account that the first term $\cos\omega_0 (t_2 - t_1)$ is a deterministic quantity, while the second term, when being statistically averaged by means of the univariate probability density $p_\theta(\theta) = 1/2\pi$ [see (4.22)], vanishes, we obtain

$$B_x(t_1, t_2) = \langle x(t_1)\,x(t_2)\rangle = \frac{1}{2} A_0^2 \cos\omega_0\tau \qquad (4.27)$$

The same result is obtained when performing the time averaging of the product $x_k(t)\,x_k(t+\tau)$ for any realization of the process.

The fact that the mean value is independent of t_1 and the autocorrelation function is independent of the position of the interval $\tau = t_2 - t_1$ on the time axis permits one to regard the process under consideration as stationary. On the other hand, the coincidence of the ensemble and time averages (for any realization) indicates that the process is ergodic. In a similar way, it is easy to show that a harmonic oscillation with a random amplitude and a random phase forms a stationary (but not ergodic) process (different realizations feature different variances).

4.2-4. Normal Random Process

The normal (Gaussian) distribution of random variables is most common in nature. The normal process is especially typical of noise in a communication channel. It is very convenient for analysis. Therefore, random processes whose distribution does not differ very much from the normal one are often replaced by the latter.

The univariate probability density of the normal process is defined by the expression

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-\bar{x})^2}{2\sigma_x^2}\right] \qquad (4.28)$$

This section deals with the stationary and ergodic normal process. Therefore, under $\bar{x}$ and σ_x^2 we may imply respectively the d-c component and the average power of the fluctuation component of a single (sufficiently long) realization of the random process.

The graphs of the probability density for some values of σ_x in the case of normal distribution are shown in Fig. 4.7. The function $p(x)$ is symmetrical with respect to the mean value. The higher the σ_x, the lower the maximum value and the flatter the curve [the area under the curve $p(x)$ is equal to unity for any value of σ_x].

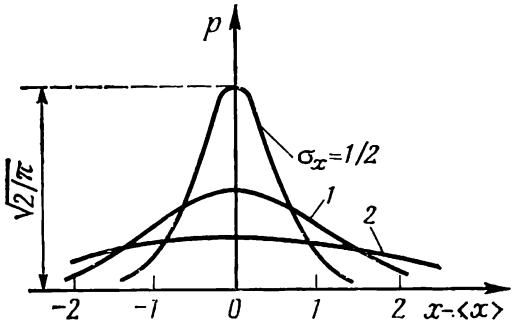


Fig. 4.7. Univariate probability density of a normal process

The function $p(x)$ can be used for finding the relative residence time of a signal $x(t)$ in a definite level interval, the ratio of its maximum value to the rms value (peak factor) and some other practically important parameters of the random signal. To illustrate, let us consider one of the realizations of normal noise, shown in Fig. 4.8*a* for $\bar{x} = 0$. This function of time corresponds to noise whose spectrum extends from the zero frequency to a certain limiting frequency. The probability that $x(t)$ falls in the interval a, b is defined by expression (4.1). Substituting (4.28) into (4.1), we have for $\bar{x} = 0$

$$P(a<x<b) = \frac{1}{\sqrt{2\pi}\sigma_x}\int_a^b e^{-x^2/2\sigma_x^2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x}\int_0^b e^{-x^2/\sigma_x^2}\,dx - \frac{1}{\sqrt{2\pi}\sigma_x}\int_0^a e^{-x^2/2\sigma_x^2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_0^{b/\sigma_x} e^{-y^2/2}\,dy - \frac{1}{\sqrt{2\pi}}\int_0^{a/\sigma_x} e^{-y^2/2}\,dy$$

$$= \Phi\left(\frac{b}{\sigma_x}\right) - \Phi\left(\frac{a}{\sigma_x}\right) \qquad (4.29)$$

The function

$$\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_0^u e^{-y^2/2}\,dy \qquad (4.30)$$

is called the probability integral. Any mathematical handbook gives tables of this function.

Substituting $b/\sigma_x = 1, 2, 3$ and, respectively, $a/\sigma_x = -1, -2$ and -3 into (4.29), we can readily find the probabilities that $x(t)$

takes a value in the bands $2\sigma_x$, $4\sigma_x$ and $6\sigma_x$ wide, which are symmetrical with respect to the axis t.

In the particular case under consideration ($|a| = b$), formula (4.29) can be simplified on account of the symmetry of the function $p(x)$ with respect to the ordinate axis (see Fig. 4.7).

Thus

$$P(-b<x<b)=2\frac{1}{\sqrt{2\pi}}\int_0^{b/\sigma_x} e^{-y^2/2}\,dy=2\Phi\left(\frac{b}{\sigma_x}\right)$$

The results of computations are summarized in Table 4.1. The last column gives the values of $1-2\,\Phi(b/\sigma_x)$. From this table it

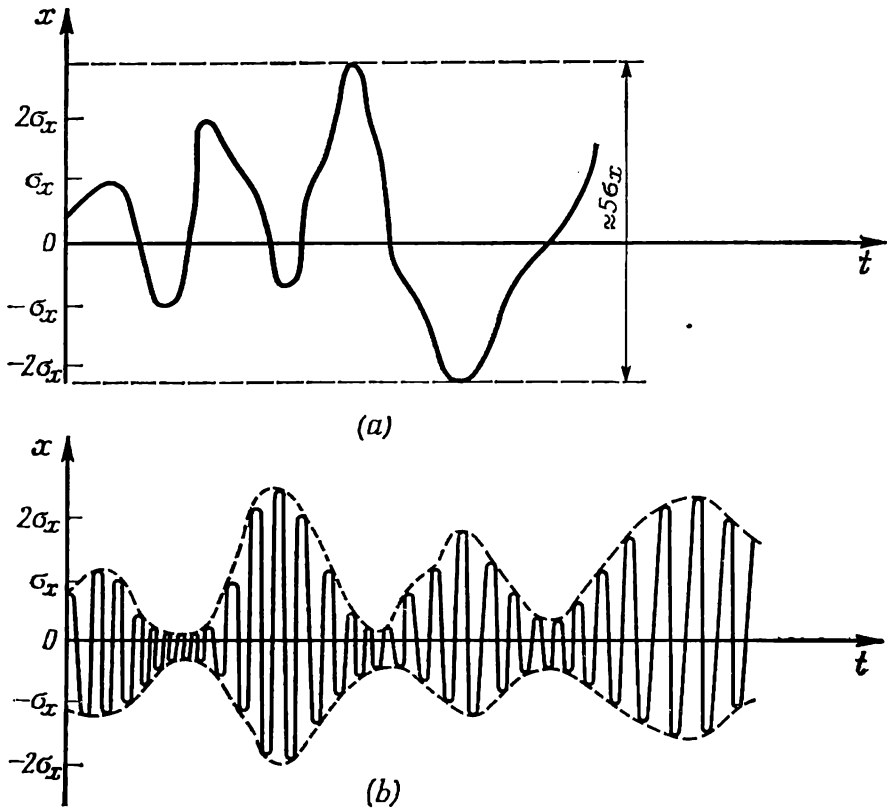


Fig. 4.8. Random functions with normal distribution but with different frequency spectra

follows that the width of the noise track (Fig. 4.8*a*) for normal noise can be taken equal to (4 to 5) σ_x. If the peaks of the function $x(t)$, whose probability is less than 1%, are taken account of, the noise peak factor can be estimated at about 3 (the peak-to-σ_x ratio). It will be recalled that the peak factor for a harmonic wave is equal to $\sqrt{2}$.

Table 4.1

Interval	Probability of residence in interval	Probability of residence outside interval
$(-\sigma_x, \sigma_x)$	$2\times 0.3413 = 0.6826$	$\approx 0.317 = 31.7\%$
$(-2\sigma_x, 2\sigma_x)$	$2\times 0.4772 = 0.9544$	$\approx 0.046 = 4.6\%$
$(-3\sigma_x, 3\sigma_x)$	$2\times 0.49865 = 0.9973$	$\approx 0.003 = 0.3\%$

The ratio of the *residence time* of $x(t)$ in a given interval to the total observation time (sufficiently large for effective averaging) can be treated as the probability that $x(t)$ falls in the given interval. Such a treatment is used as a basis for designing various instruments for finding experimentally the univariate probability density of a random process. It can be noted that the above data on the probability distribution give no idea of the behaviour of the function $x(t)$ *in time*. Figure 4.8*b* shows a realization of normal noise, whose spectrum is concentrated in a narrow frequency band with a centre frequency ω_0. As to the probability density $p(x)$ and consequently the values of quantities $\bar{x}$ and σ_x, this noise does not differ from the low-frequency noise shown in Fig. 4.8*a*.

To describe the time characteristics of the function $x(t)$, it is necessary to use the bivariate probability density which makes it possible to find the autocorrelation function. Another method consists in finding the power spectrum of a random process. This is discussed in the next section.

4.3. POWER SPECTRAL DENSITY OF A RANDOM PROCESS

Implying an ensemble of realizations under a random process, one should remember that the realizations of different forms have different spectral characteristics. The averaging of the complex spectral density, introduced in Sec. 2.6 or in Sec. 2.12, over all realizations results in zero spectrum of the process (at $\bar{x} = 0$), because of the randomness and independence of the phases of the spectral components in the different realizations. However, we may introduce the concept of the *spectral density of the mean square* of a random function, since the mean square value is independent of the relation between the phases of the harmonics being added together. If a random function $x(t)$ represents a voltage or current, the mean square of this function can be regarded as the average power dissipated in a resistor of 1 ohm. This power is distributed over a frequency band depending on the mechanism of formation of the random process. The spectral density of the average power is the *average power per hertz at a given frequen-*

cy ω. This spectral density, $W(\omega)$, will further be called the *energy spectrum of the function* $x(t)$. The meaning of this term can be inferred from the dimension of the function $W(\omega)$ which is the ratio of power to frequency band:

$$W(\omega)=\left[\frac{Power}{Frequency\ Band}\right]=[Power\times Time]=[Energy]$$

The energy spectrum can be found if the formation mechanism of the random process is known. As applied to noise associated with the atomism of matter and electricity, this problem will be discussed in Sec. 7.2. Here we shall restrict ourselves to some general definitions.

Having selected from the ensemble a realization $x_k(t)$ within a finite time interval T, we can find the spectral density $\mathrm{X}_{kT}(\omega)$ by means of ordinary Fourier transform. Then the energy of the realization in this interval can be calculated by formula (2.66):

$$E_{kT}=\int_{-T/2}^{T/2} x_{kT}^2(t)\,dt=\frac{1}{2\pi}\int_{-\infty}^{\infty}|\mathrm{X}_{kT}(\omega)|^2\,d\omega \tag{4.31}$$

Dividing this energy by T, we obtain the average power of the kth realization in the interval T:

$$\overline{x_{kT}^2(t)}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|\mathrm{X}_{kT}(\omega)|^2}{T}\,d\omega$$

The energy E_{kT} increases with an increase in T, however, the ratio E_{kT}/T tends to a certain limit. To extend our reasoning to a stationary random process, it is necessary to let the duration T of each realization go to infinity. But in this case a formal difficulty emerges, since the Fourier transform exists only for signals of finite energy (the conditions of the absolute integrability of the signal). This difficulty can be overcome by various methods [8]. In the given case we shall only stipulate that T is arbitrarily large but *finite*. Subject to this condition, we shall write the last expression for the average power of kth realization in the form

$$\overline{x_k^2(t)}=\frac{1}{2\pi}\int_{-\infty}^{\infty}W_k(\omega)\,d\omega \tag{4.32}$$

where

$$W_k(\omega)=|\mathrm{X}_{kT}(\omega)|^2/T \tag{4.33}$$

is the *average power spectral density* of the kth realization being studied (at a sufficiently large T).

In the general case, the quantity $W_k(\omega)$ must be averaged over the ensemble of realizations. Restricting ourselves here to a *stationary* and *ergodic* process, we may consider that the function $W_k(\omega)$, found by the averaging of a single realization, characterizes the process as a whole. Omitting the subscript k, we get a final expression for the average power of a random process:

$$\overline{x^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\omega)\, d\omega \tag{4.34}$$

If we consider a random process with a nonzero average value of $x(t)$, the energy spectrum should be represented in the form

$$W_x(\omega) = [\overline{x(t)}]^2 \times 2\pi\delta(\omega) + W_{\sim}(\omega) \tag{4.35}$$

where $W_{\sim}(\omega)$ is the continuous portion of the spectrum, corresponding to the fluctuation component x, and $\delta(\omega)$ is the delta function.

On integrating with respect to $f = \omega/2\pi$, the first term on the right-hand side gives $[\overline{x(t)}]^2$, i.e., the power of the d-c component, while the second term gives the power of the fluctuation component, i.e., the variance

$$\sigma_x^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_{\sim}(\omega)\, d\omega \tag{4.36}$$

For the process with zero mean value we have

$$\overline{x^2(t)} = \sigma_x^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_x(\omega)\, d\omega \tag{4.37}$$

From the definition of the energy spectrum (4.33) it is clear that $W(\omega)$ is an even and nonnegative function of ω.

4.4. RELATIONSHIP BETWEEN THE ENERGY SPECTRUM AND THE AUTOCORRELATION FUNCTION OF A RANDOM PROCESS

The slower the change of $x(t)$ in time, the narrower the energy spectrum. On the other hand, the rate of change of $x(t)$ determines the run of the autocorrelation function. It is clear that $W_x(\omega)$ and $B_x(\tau)$ are closely related.

The Wiener-Khintchine theorem states that $B_x(\tau)$ and $W_x(\omega)$ are each other's Fourier transforms:

$$W_x(\omega) = \int_{-\infty}^{\infty} B_x(\tau)\, e^{-i\omega\tau}\, d\tau \tag{4.38}$$

$$B_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_x(\omega)\, e^{i\omega\tau}\, d\omega \tag{4.39}$$

From these expressions follows a property similar to the properties of Fourier transforms given in Chap. 2 for deterministic signals: *the wider the energy spectrum of a random process, the shorter the correlation time, and respectively, the longer the correlation time, the narrower the spectrum of the process.*

Of special interest is white noise whose energy spectrum is uniform at all frequencies $-\infty < \omega < \infty$.

Substituting $W_x(\omega) = W_0 = \text{const}$ into expression (4.39), we obtain [see (2.93)]

$$B_x(\tau) = W_0 \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} d\omega = W_0 \delta(\tau) \tag{4.40}$$

where $\delta(\tau)$ is the delta function.

For white noise of infinite and uniform spectrum, the autocorrelation function is zero for all τ except $\tau = 0$ at which $B_x(0) = \infty$.

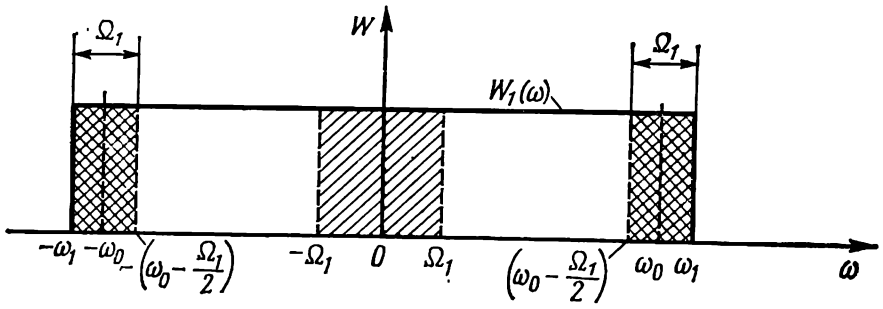


Fig. 4.9. Wide-band and narrow-band energy spectra (examples 1, 2 and 3)

Such a noise having a needle structure with infinitely thin random spikes, is sometimes called delta-correlated process. The variance of white noise is infinitely large.

Let us illustrate the above relations with the following examples.

1. Let the following parameters of noise voltage (a normal stationary process with zero mean) be given: the root-mean square value $v_{rms} = 2\text{V}$, the energy spectrum $W_1(\omega)$ is uniform in the frequency range from $-f_1$ to $f_1 = 10$ MHz (solid line in Fig. 4.9).

Noise with such a spectrum is usually called *wide-band.* In the given case

$$W_1(\omega) = v_{rms}^2/2f_1 = (2)^2/2 \times 10^7 = 2 \times 10^{-7}\ \text{V}^2/\text{Hz}$$

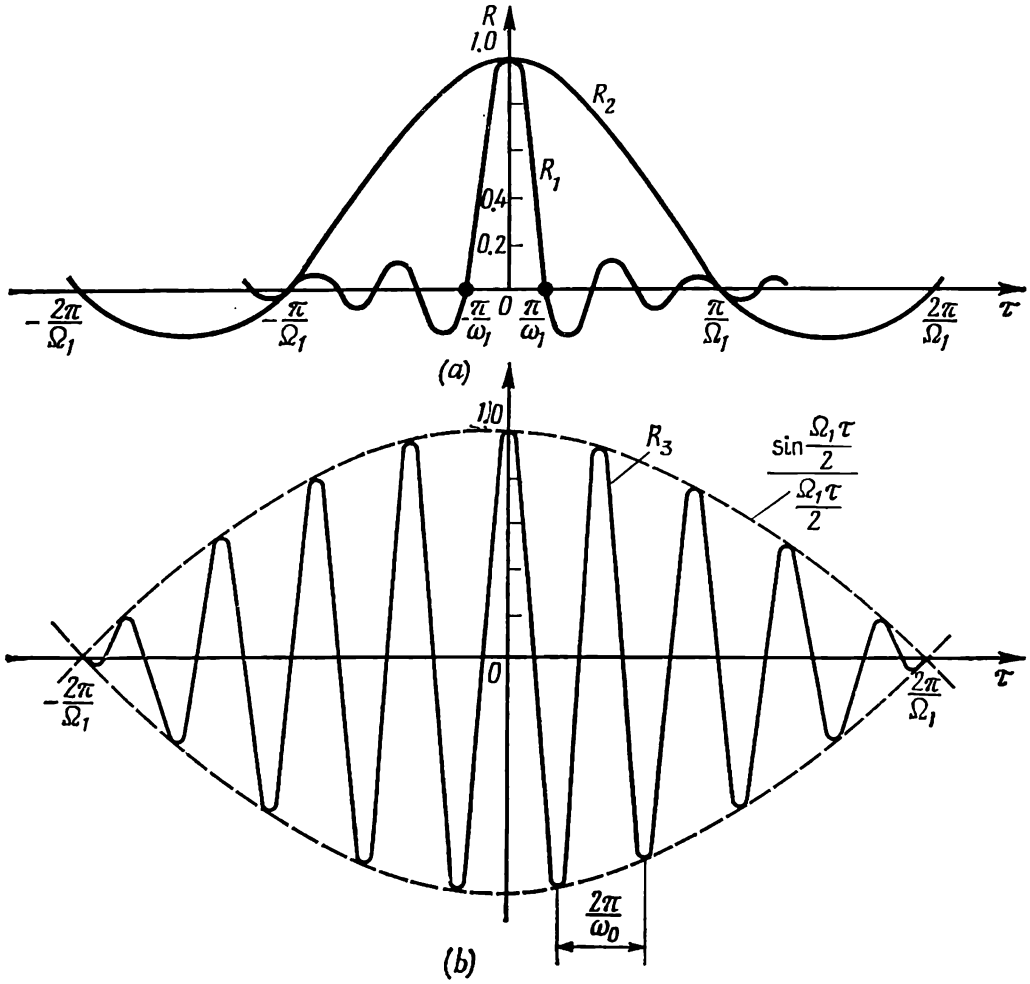


Fig. 4.10. Normalized autocorrelation function of a random process whose energy spectrum is uniform within a band of (*a*) $|\omega| \leqslant \omega_1$ and $|\omega| \leqslant \Omega_1 = \omega_1/5$ and (*b*) $\omega_0 - \Omega_1/2 \leqslant \omega \leqslant \omega_0 + \Omega_1/2$

The autocorrelation function of the process [see (4.39)] is

$$B_1(\tau) = \frac{1}{2\pi}\int_{-\omega_1}^{\omega_1} W_1(\omega)\, e^{i\omega\tau}\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\omega_1}^{\omega_1} W_1(\omega)\cos\omega\tau\, d\omega = 2\times 10^{-7}\,\frac{1}{2\pi}\,\frac{2\sin\omega_1\tau}{\tau}$$

$$= 2\times 10^{-7}\, 2f_1\,\frac{\sin\omega_1\tau}{\omega_1\tau} = \sigma_1^2\,\frac{\sin\omega_1\tau}{\omega_1\tau} \qquad (4.41)$$

The variance is given by

$$\sigma_1^2 = v_{rms}^2 = B_1(0) = 4 \text{ V}^2$$

The normalized autocorrelation function (Fig. 4.10*a*) is

$$R_1(\tau) = B_1(\tau)/\sigma_1^2 = \sin \omega_1\tau/\omega_1\tau \tag{4.42}$$

2. Let us isolate a frequency band from $\omega = -\Omega_1 = -2\pi F_1$ to $\omega = \Omega_1 = 2\pi F_1$ (hatched area in Fig. 4.9) from the spectrum of the original noise and find $B_2(\tau)$, $R_2(\tau)$, and σ_2^2 corresponding to this band-limited noise.

At $F_1 = 2$ MHz we obtain

$$\sigma_2^2 = 2F_1W_1(\omega) = 2 \times 2 \times 10^6 \times 2 \times 10^{-7} = 0.8 \text{ V}^2$$

$$B_2(\tau) = 2 \times 10^{-7} \times 2F_1 \frac{\sin \Omega_1\tau}{\Omega_1\tau} = 0.8 \frac{\sin \Omega_1\tau}{\Omega_1\tau}$$

$$R_2(\tau) = B_2(\tau)/\sigma_2^2 = (\sin \Omega_1\tau)/\Omega_1\tau$$

The narrowing of the spectrum has resulted in the expansion of the curve $R_2(\tau)$ along the τ-axis (Fig. 4.10*a*). The correlation time has increased by a factor of $f_1/F_1 = 5$.

3. Now let us find similar characteristics for noise whose spectrum is shown in Fig. 4.9 as crosshatched areas. This case differs from the preceding one in the position of the spectral band on the frequency axis. Noise with such a spectrum is called *narrow-band* (at $\Omega_1/\omega_0 \ll 1$).

The variance of this noise apparently does not differ from σ_2^2.

The autocorrelation function is

$$B_3(\tau) = \frac{1}{2\pi} \int_{-(\omega_0+\Omega_1/2)}^{-(\omega_0-\Omega_1/2)} W_1(\omega)\, e^{\omega\tau}\, d\omega + \frac{1}{2\pi} \int_{\omega_0-\Omega_1/2}^{\omega_0+\Omega_1/2} W_1(\omega)\, e^{i\omega\tau}\, d\omega$$

$$= \frac{1}{\pi} \int_{\omega_0-\Omega_1/2}^{\omega_0+\Omega_1/2} W_1(\omega) \cos \omega\tau\, d\omega$$

$$= 2 \times 10^{-7} \frac{1}{\pi} \left[\frac{\sin(\omega_0+\Omega_1/2)\tau}{\tau} - \frac{\sin(\omega_0-\Omega_1/2)\tau}{\tau} \right]$$

$$= 2 \times 10^{-7} \frac{1}{\pi\tau} \times 2 \sin \frac{\Omega_1\tau}{2} \cos \omega_0\tau$$

$$= 2 \times 10^{-7} \times 2F_1 \frac{\sin \Omega_1\tau/2}{\Omega_1\tau/2} \cos \omega_0\tau \tag{4.43}$$

The normalized autocorrelation function (Fig. 4.10*b*) is given by

$$R_3(\tau) = \frac{\sin(\Omega_1\tau/2)}{\Omega_1\tau/2} \cos \omega_0\tau \tag{4.44}$$

The shape of the envelope of the function $R_3(\tau)$ (dashed line) is similar to that of the function $R_2(\tau)$, but the correlation time of the former is twice that of the latter. The high-frequency oscillation inscribed in the envelope of the function $R_2(\tau)$ has a frequency ω_0 equal to the centre frequency of noise (see Fig. 4.9).

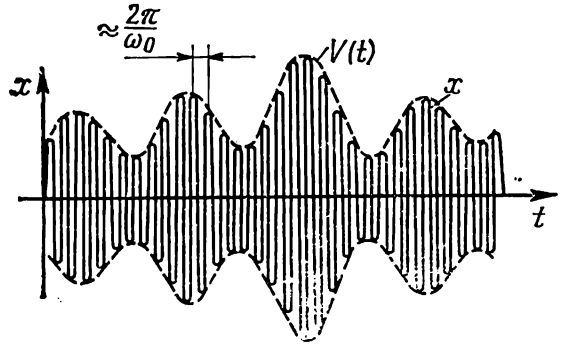


Fig. 4.11. General view of a realization of the random process whose autocorrelation function is shown in Fig. 4.10*b* (the scales along the axes t and τ are different)

The graph of the normalized autocorrelation function shown in Fig. 4.10*b* gives an idea of the nature of a narrow-band noise signal. The oscillations of the autocorrelation function with a frequency ω_0 indicate to the fact that the instantaneous value of the noise signal, *on the average*, changes at a frequency ω_0. It will be recalled that the autocorrelation function of a harmonic wave is also a harmonic function of the same frequency (see Sec. 2.16). On the other hand, the change in the envelope of the autocorrelation function according to $\frac{\sin(\Omega_1\tau/2)}{\Omega_1\tau/2}$ points to the fact that the envelope of the noise oscillation, which is a random variable, changes in time relatively slowly, like a function of time, whose spectrum is limited by the highest frequency Ω_1. A general view of a noise signal corresponding to the autocorrelation function (4.44) is presented in Fig. 4.11 (not to scale along the abscissa axis).

Thus a narrow-band noise signal is like a high-frequency wave whose amplitude and phase change slowly as compared with the frequency ω_0:

$$x(t) = V(t)\cos[\omega_0 t + \theta(t)] \qquad (4.45)$$

where ω_0 is the centre frequency of the noise spectrum.

It should be noted that all the parameters of this oscillation (amplitude, phase, and frequency) are random functions of time. The statistical characteristics of these parameters are discussed in Sec. 4.6.

4.5. CROSS-CORRELATION FUNCTION AND CROSS-SPECTRAL DENSITY OF TWO RANDOM PROCESSES

The degree of association between two stationary processes, $x(t)$ and $y(t)$, is estimated by means of cross-correlation functions defined by the expressions *

$$\begin{aligned} B_{xy}(\tau) &= \langle x(t)\, y(t+\tau)\rangle \\ B_{yx}(\tau) &= \langle y(t)\, x(t+\tau)\rangle \end{aligned} \tag{4.46}$$

This section deals with ergodic processes, therefore (4.46) can be replaced by time averages

$$B_{xy}(\tau) = \overline{x(t)\, y(t+\tau)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, y(t+\tau)\, dt \tag{4.47}$$

$$B_{yx}(\tau) = \overline{y(t)\, x(t+\tau)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} y(t)\, x(t+\tau)\, dt \tag{4.48}$$

As in the case of deterministic signals, the cross-correlation function does not change if the time shift τ of one of the functions $x(t)$ or $y(t)$ is replaced by the reverse shift of the other function. Therefore, the following equations may be written:

$$B_{xy}(\tau) = \overline{x(t)\, y(t+\tau)} = \overline{x(t-\tau)\, y(t)} \tag{4.49}$$

$$B_{yx}(\tau) = \overline{y(t)\, x(t+\tau)} = \overline{y(t-\tau)\, x(t)} \tag{4.50}$$

From the last expressions stem the following relations between $B_{xy}(\tau)$ and $B_{yx}(\tau)$:

$$B_{xy}(\tau) = B_{yx}(-\tau), \quad B_{yx}(\tau) = B_{xy}(-\tau) \tag{4.51}$$

Relations (4.49) through (4.51) should not be confused with the conditions for the *even parity* of the functions. Each of the functions $B_{xy}(\tau)$ and $B_{yx}(\tau)$ is not necessarily even with respect to τ (see Sec. 2.16).

As a result, the correlation between the values of the functions $x(t)$ and $y(t)$ at two different instants of time separated by an interval τ is specified by the coherency matrix

$$B(\tau) = \begin{bmatrix} B_{xx}(\tau) & B_{xy}(\tau) \\ B_{yx}(\tau) & B_{yy}(\tau) \end{bmatrix} \tag{4.52}$$

where $B_{xx}(\tau)$ and $B_{yy}(\tau)$ are the autocorrelation functions of the processes $x(t)$ and $y(t)$, respectively.

* It is implied that the relations between the processes $x(t)$ and $y(t)$ are stationary, as well as the processes themselves.

For example, let us consider a sum of two ergodic processes, $x(t)$ and $y(t)$, with zero means ($\bar{x} = \bar{y} = 0$) and determine the autocorrelation function of the random process $s(t) = x(t) + y(t)$ (provided that the cross-correlation functions are stationary).

Using formula (4.16) and taking into account (4.49) and (4.50), we obtain

$$B_s(\tau) = \overline{s(t)\, s(t+\tau)} = \overline{[x(t)+y(t)]\,[x(t+\tau)+y(t+\tau)]}$$
$$= \overline{x(t)\, x(t+\tau)} + \overline{x(t)\, y(t+\tau)} + \overline{y(t)\, x(t+\tau)} + \overline{y(t)\, y(t+\tau)}$$
$$= B_{xx}(\tau) + B_{xy}(\tau) + B_{yx}(\tau) + B_{yy}(\tau) \qquad (4.53)$$

When $\tau = 0$, $B_{xx}(0) = \sigma_x^2$ and $B_{yy}(0) = \sigma_y^2$, while $B_{xy}(0) = B_{yx}(0)$. Consequently,

$$\sigma_s^2 = B_s(0) = \sigma_x^2 + \sigma_y^2 + B_{xy}(0) + B_{yx}(0) = \sigma_x^2 + \sigma_y^2 + 2B_{xy}(0) \qquad (4.53')$$

If the processes $x(t)$ and $y(t)$ are statistically independent, the variance (average power) of the sum is $\sigma_s^2 = \sigma_x^2 + \sigma_y^2$.

Otherwise, depending on the sign of $B_{xy}(0)$, the power of the process $s(t)$ may be higher or lower than the sum of the variances σ_x^2 and σ_y^2.

For the difference $s(t) = x(t) - y(t)$ we obtain an expression similar to (4.53'), the only thing necessary being the reversal of

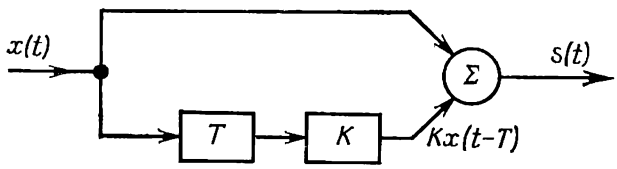


Fig. 4.12. Determining the autocorrelation function of two random processes with identical energy spectra

sign before the term $2B_{xy}$. With the processes $x(t)$ and $y(t)$ being statistically independent, the variance of the process $s(t)$ is $\sigma_s^2 = \sigma_x^2 + \sigma_y^2$, as in the case of summation.

In practice, one often encounters a case of summing a process $x(t)$ with a process $Kx(t-T)$, i.e., with the same process delayed by a time T and amplified by a factor of K (Fig. 4.12).

Let us set up matrix (4.52) for the processes $x(t)$ and $y = Kx(t-T)$. Using the symbols of (4.52), we obtain

$$B_{xx}(\tau) = B_x(\tau)$$
$$B_{xy}(\tau) = \overline{x(t)\, y(t+\tau)} = \overline{Kx(t)\, x(t-T+\tau)} = KB_x(\tau - T)$$
$$B_{yx}(\tau) = \overline{y(t)\, x(t+\tau)} = \overline{Kx(t-T)\, x(t+\tau)} = KB_x(\tau + T)$$
$$B_{yy}(\tau) = B_y(\tau) = \overline{y(t)\, y(t+\tau)} = \overline{K^2 x(t-T)\, x(t-T+\tau)} = K^2 B_x(\tau)$$

Thus, the coherency matrix of the processes $x(t)$ and $y(t) = Kx(t-T)$ assumes the form

$$B(\tau) = \begin{bmatrix} B_x(\tau) & KB_x(\tau - T) \\ KB_x(\tau + T) & K^2 B_x(\tau) \end{bmatrix}$$

Now let us find the autocorrelation function of the process $s(t) = x(t) + y(t)$ at the adder output (Fig. 4.12). Substituting the elements of the matrix $B(\tau)$ into (4.53), we get

$$B_s(\tau) = B_x(\tau) + KB_x(\tau - T) + KB_x(\tau + T) + K^2 B_x^2(\tau)$$

Setting $\tau = 0$, we find the variance of the process:

$$\sigma_s^2 = \sigma_x^2 + KB_x(-T) + KB_x(T) + K^2\sigma_x^2 = (1 + K^2)\,\sigma_x^2 + 2KB_x(T) = \sigma_x^2 [1 + K^2 + 2KR_x(T)],$$

where $R_x(T) = B_x(T)/\sigma_x^2$ is the normalized autocorrelation function of the process $x(t)$ (it will be recalled that in the given example $\overline{x(t)} = 0$).

If the adder is replaced by a subtractor, the sign before the term $2KR_x(T)$ must be reversed.

If the delay T is much longer than the correlation time of the process $x(t)$, then $R_x(T) \to 0$ and $\sigma_s^2 = \sigma_x^2(1 + K^2)$.

Let us now apply the Wiener-Khintchine relation (4.38) to $B_s(\tau)$

$$W_s(\omega) = \int_{-\infty}^{\infty} B_s(\tau)\, e^{-i\omega\tau}\, d\tau = W_x(\omega) + W_y(\omega) + W_{xy}(\omega) + W_{yx}(\omega) \tag{4.54}$$

In this expression

$$W_{xy}(\omega) = \int_{-\infty}^{\infty} B_{xy}(\tau)\, e^{-i\omega\tau}\, d\tau, \quad W_{yx}(\omega) = \int_{-\infty}^{\infty} B_{yx}(\tau)\, e^{-i\omega\tau}\, d\tau \tag{4.55}$$

have the meaning of the cross-spectral densities of the random processes $x(t)$ and $y(t)$.

The inverse Fourier transforms of $W_{xy}(\omega)$ and $W_{yx}(\omega)$ assume the form

$$\left.\begin{aligned} B_{xy}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} W_{xy}(\omega)\, e^{i\omega\tau}\, d\omega \\ B_{yx}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} W_{yx}(\omega)\, e^{i\omega\tau}\, d\omega \end{aligned}\right\} \tag{4.56}$$

In contrast to the energy spectrum $W_x(\omega)$, or $W_y(\omega)$, which is a real function of ω and cannot assume negative values, the cross-spectral densities $W_{xy}(\omega)$ and $W_{yx}(\omega)$ may be complex functions. This happens when the functions $B_{xy}(\tau)$ and $B_{yx}(\tau)$ are odd with respect to τ. Substitution of (4.51) into (4.55) gives

$$W_{xy}(\omega) = W_{yx}^*(\omega) \tag{4.57}$$

from which it follows that

$$W_{xy}(\omega) + W_{yx}(\omega) = 2\mathrm{Re}\,[W_{xy}(\omega)] = 2\mathrm{Re}\,[W_{yx}(\omega)] \tag{4.58}$$

Thus, expression (4.54) may be written in the form

$$W_s(\omega) = W_x(\omega) + W_y(\omega) + 2\mathrm{Re}\,[W_{xy}(\omega)] \tag{4.59}$$

This expression clarifies the physical meaning of the cross-spectral density $W_{xy}(\omega)$. If the random processes $x(t)$ and $y(t)$ are statistically independent, $W_{xy}(\omega) = 0$ and the energy spectrum of the sum $s(t) = x(t) + y(t)$ is equal to the sum of the energy spectra $W_x(\omega)$ and $W_y(\omega)$, and consequently, the power of the process $s(t)$ is equal to the sum of the powers of the processes $x(t)$ and $y(t)$.

If the real part of the cross-spectral density is positive, $W_s(\omega) > W_x(\omega) + W_y(\omega)$, which means that the correlation between the processes in this case leads to an increase in the average power of the resultant process ($\sigma_s^2 > \sigma_x^2 + \sigma_y^2$). It is clear that with a negative real part of $W_{xy}(\omega)$, the average power of the resultant process is less than $\sigma_x^2 + \sigma_y^2$.

If $\sigma_s^2 = \sigma_x^2 + \sigma_y^2$, the processes $x(t)$ and $y(t)$ are incoherent and additive (see Sec. 2.18).

4.6. NARROW-BAND RANDOM PROCESS

A brief description of the properties of normal noise formed from white noise by isolating a comparatively narrow frequency band was given in Sec. 4.4. It was noted that each realization of such a random process is almost a harmonic wave

$$x(t) = A(t)\cos[\omega_0 t + \theta(t)] = A(t)\cos\psi(t) \tag{4.60}$$

whose parameters — the envelope $A(t)$, phase $\theta(t)$, and frequency $\omega(t)$ — are slowly varying random functions of time.

When noise is represented in the form (4.60), it is assumed that the envelope $A(t)$ obeys the expression

$$A(t) = \sqrt{x^2(t) + y^2(t)} \tag{4.61}$$

where $y(t)$ is the Hilbert transform (harmonic conjugate) of the original function $x(t)$ (see Sec. 3.9), and that ω_0 does not contain any terms linearly dependent on t.

Further consideration is based on the premise that the energy spectrum of noise $x(t)$ is concentrated in a narrow (compared with ω_0) frequency band, the function $W_x(\omega)$ in this band being symmetrical with respect to the point ω_0 (Fig. 4.13*a*).

Let us consider a stationary ergodic process obeying a normal law of probability distribution. It should be pointed out here that this

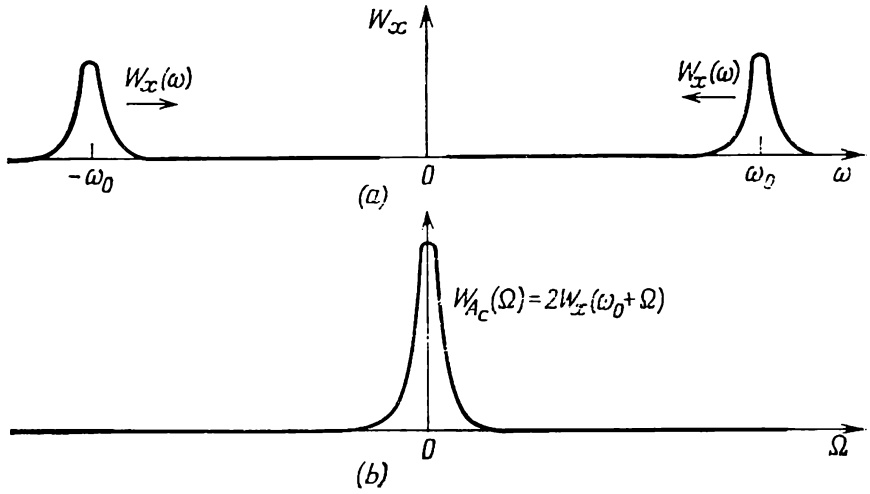


Fig. 4.13. Energy spectra of (*a*) a narrow-band process with a centre frequency ω_0 and (*b*) the cosine component of the complex envelope

distribution characterizes the physical oscillation $x(t)$, i.e., the instantaneous value of the oscillation (at any instant of time t), whereas the oscillation parameters $A(t)$, $\theta(t)$, and $\omega(t) = d\psi/dt$ obey distribution laws other than normal *. To fully describe the properties of a narrow-band process, one must know the distribution laws and correlation functions of all the parameters of the oscillation.

4.6-1. The Envelope

Let a high-frequency oscillation $x(t)$, defined by expression (4.60), be represented in the form of a sum of two quadrature components:

$$x(t) = A(t)\cos\theta(t)\cos\omega_0 t - A(t)\sin\theta(t)\sin\omega_0 t$$
$$= A_c(t)\cos\omega_0 t - A_s(t)\sin\omega_0 t \qquad (4.60')$$

As in Sec. 3.5,

$$A_c(t) = A(t)\cos\theta \text{ and } A_s(t) = A(t)\sin\theta \qquad (4.62)$$

* This stems from the nonlinear character of the dependence of the parameters A, θ, and ω on x and y.

are the amplitudes of the cosine and sine components of the oscillation $x(t)$, respectively, and

$$A(t) = \sqrt{A_c^2(t) + A_s^2(t)}, \quad \theta(t) = -\arctan A_s/A_c \tag{4.63}$$

To find the probability densities $p_A(A)$ and $p_\theta(\theta)$, one must know the respective densities $p(A_c)$ and $p(A_s)$, and also the joint probability density $p(A_c, A_s)$.

The densities $p(A_c)$ and $p(A_s)$ can be found by comparing the random function $A_c(t)$ [or $A_s(t)$] with the function $x(t)$:

$$x(t) = A(t)\cos[\omega_0 t + \theta(t)];$$
$$A_c(t) = A(t)\cos\theta(t)$$

The distinction of $A_c(t)$ from $x(t)$ consists in the absence of the term $\omega_0 t$ in the argument of the cosine. As in the case of a deterministic oscillation, this means a shift of each realization of the random process by an amount ω_0 (in the direction towards the zero frequency, the *spectrum structure being preserved*). In this case the distribution law of the random function $x(t)$ remains unchanged. Therefore, if the process $x(t)$ is normal, the process $A_c(t)$ is also normal (both processes have zero mean).

The energy spectrum $W_{A_c}(\Omega)$ of the random function $A_c(t)$ (Fig. 4.13*b*) can be obtained from the energy spectrum of the function $x(t)$ by shifting the left-hand lobe of the spectrum $W_x(\omega)$ by an amount ω_0 and the right-hand lobe, by an amount $-\omega_0$ (Fig. 4.13*a*).

As a result, we have the energy spectrum

$$W_{Ac}(\Omega) = 2W_x(\omega_0 + \Omega) \tag{4.64}$$

concentrating near the zero frequency. The factor 2 takes into account the summation of the power of both lobes of $W_x(\omega)$*.

By similar reasoning, for the random process $A_s(t)$ and its energy spectrum, we have

$$W_{As}(\Omega) = 2W_x(\omega_0 + \Omega) = W_{Ac}(\Omega)$$

From this expression and Fig. 4.13*b* it follows that the area under the curve $W_x(\omega)$ (both lobes) is equal to the area under the curve $W_{A_c}(\Omega)$ [or $W_{A_s}(\Omega)$]. Consequently, the variances of the random functions $A_c(t)$, $A_s(t)$, and $x(t)$ are the same:

$$\sigma_{Ac}^2 = \sigma_A^2 = \sigma_x^2$$

* In the case of a deterministic amplitude-modulated oscillation (Fig. 3.9), when transforming from the spectrum $S_a(\omega)$ to the spectrum $S_A(\omega)$, the spectral density of current (or voltage) is doubled, this resulting in the quadruplication of the spectral energy density proportional to $S_A^2(\omega)$. In the given case the power is only doubled because of an incoherent addition of the energy spectra of both lobes of $W_x(\omega)$.

Taking into account the first expression of (4.63), from which it follows that $A^2(t) = A_c^2(t) + A_s^2(t)$, we obtain the following expression for the mean square of the envelope

$$\langle A^2 \rangle = \overline{A^2(t)} = \sigma_{A_c}^2 + \sigma_{A_s}^2 = 2\sigma_x^2 \tag{4.65}$$

Thus, the univariate probability densities of random functions $A_c(t)$ and $A_s(t)$ can be defined by the expressions

$$p(A_c) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{A_c^2}{2\sigma_x^2}\right),$$

$$p(A_s) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{A_s^2}{2\sigma_x^2}\right) \tag{4.66}$$

Furthermore, the cross-correlation between the functions $A_c(t)$ and $A_s(t)$ is zero at $\tau = 0$. Indeed, squaring expression (4.60′) and averaging it over the ensemble, we obtain.

$$\langle x^2(t) \rangle = \langle [A_c(t) \cos \omega_0 t - A_s(t) \sin \omega_0 t]^2 \rangle$$

$$= \langle A_c^2(t) \rangle \cos^2 \omega_0 t + \langle A_s^2(t) \rangle \sin^2 \omega_0 t - 2 \langle A_c(t) A_s(t) \rangle \sin \omega_0 t \cos \omega_0 t$$

But the left-hand side of this expression is equal to $B_x(0) = \sigma_x^2$, and besides, $\langle A_c^2(t) \rangle = \langle A_s^2(t) \rangle = \sigma_x^2 = B_{A_c}(0) = B_{A_s}(0)$, and $\langle A_c(t) A_s(t) \rangle = B_{A_c A_s}(0)$ is the cross-correlation function of the random processes $A_c(t)$ and $A_s(t)$ at $\tau = 0$. Therefore, the preceding equation is reduced to the form

$$B_x(0) = B_{A_c}(0) - B_{A_c A_s}(0) \sin 2\omega_0 t \tag{4.67}$$

from which it follows that $B_{A_c A_s}(0) = 0$ [since the processes $x(t)$ and $A_c(t)$ are stationary, the equality (4.67) must hold at any instant of time].

Thus, $A_c(t)$ and $A_s(t)$, taken at the same instant of time, are statistically independent*. Therefore, the joint probability density $p(A_c, A_s)$ may be defined by the expression

$$p(A_c, A_s) = p(A_c)\, p(A_s) = \frac{1}{2\pi\sigma_x^2} \exp\left(\frac{-A_c^2 - A_s^2}{2\sigma_x^2}\right)$$

$$= \frac{1}{2\pi\sigma_x^2} \exp\left(-\frac{A^2}{2\sigma_x^2}\right) \tag{4.68}$$

The probability that the end of the vector $A(t)$ lies in the elementary rectangle $dA_c dA_s$ (Fig. 4.14) is equal to the product of the pro-

* This also follows from relation (4.65) showing that the mean square of the envelope $A(t)$, i.e., σ_A^2 is an additive sum of the mean squares of the function $A_c(t)$ and $A_s(t)$.

bability of residence of A_c in the interval dA_c and that of A_s in the interval dA_s:

$$p(A_c)\,dA_c\,p(A_s)\,dA_s = \frac{1}{2\pi\sigma_x^2}\exp\left(-\frac{A^2}{2\sigma_x^2}\right)dA_c dA_s$$

Transformed from rectangular to polar coordinates, the area of the hatched element in Fig. 4.15 is $Ad\theta dA$, and the probability of

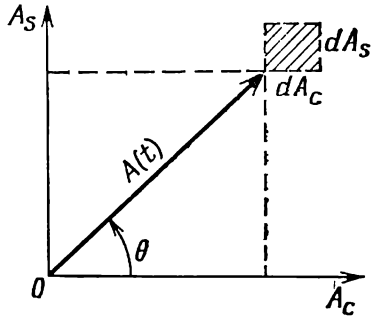


Fig. 4.14. Defining the bivarite probability density of the quadrature components of the complex envelope of a random process

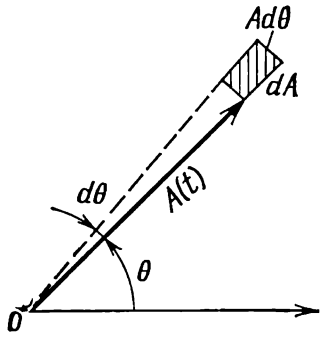


Fig. 4.15. Defining the bivariate probability densities of the modulus and argument of the complex envelope

residence of the end of the vector in this element is equal to

$$\frac{1}{2\pi\sigma_x^2}\exp\left(-\frac{A^2}{2\sigma_x^2}\right)Ad\theta dA$$

From this expression it follows that the bivariate probability density

$$p(A,\theta) = \frac{A}{2\pi\sigma_x^2}\exp\left(-\frac{A^2}{2\sigma_x^2}\right) \tag{4.69}$$

Integrating with respect to θ, we obtain the univariate probability density

$$p_A(A) = \int_{-\pi}^{\pi} p(A,\theta)\,d\theta = \frac{A}{\sigma_x^2}\exp\left(-\frac{A^2}{2\sigma_x^2}\right),$$

$$0 < A < \infty \tag{4.70}$$

The substantiation of the integral limits is given in the next paragraph of this section.

The distribution of the envelope, characterized by probability density (4.70) is called *Rayleigh distribution* (Fig. 4.16). The maximum value of the function $p_A(A)$ occurs at $A = \sigma_x$. This means that $A = \sigma_x$ is the most probable value of the envelope.

The mean value (mathematical expectation) of the envelope is given by

$$\langle A\rangle = \int_0^\infty A p_A(A)\,dA = \frac{1}{\sigma_x^2}\int_0^\infty A^2 \exp\left(-\frac{A^2}{2\sigma_x^2}\right) dA = \sqrt{\frac{\pi}{2}}\,\sigma_x \qquad (4.71)$$

In a similar way, the mean square of the envelope is

$$\langle A\rangle^2 = \int_0^\infty A^2 p_A(A)\,dA = \frac{1}{\sigma_x^2}\int_0^\infty A^3 \exp\left(-\frac{A^2}{2\sigma_x^2}\right) dA = 2\sigma_x^2 \qquad (4.72)$$

This result coincides with (4.65).

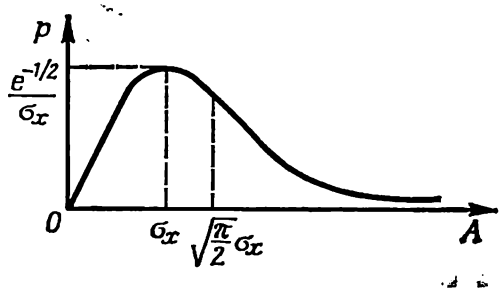


Fig. 4.16. Probability density of the Rayleigh distribution

Thus, the average power of the envelope is twice the variance of noise. This is similar to the relation between the square of the amplitude A_0 and the average power of a harmonic wave $a(t) = A_0 \cos\omega_0 t$, which is equal to $\overline{a^2(t)} = \frac{1}{2}A_0^2$.

The probability that the envelope $A(t)$ exceeds a certain specified level C is defined by the formula

$$P(A > C) = \int_C^\infty p_A(A)\,dA = \frac{1}{\sigma_x^2}\int_C^\infty A \exp\left(-\frac{A^2}{2\sigma_x^2}\right) dA = \exp\left(-C^2/2\sigma_x^2\right) \qquad (4.73)$$

and the probability that the envelope $A(t)$ falls below the level C, by the formula

$$P(A < C) = 1 - \exp\left(-C^2/2\sigma_x^2\right) \qquad (4.74)$$

From these formulas it is obvious that already at $C = 2\sigma_x^2$ the probability of exceeding the level C is only about 1%. Thus, we may consider that the width of the actually observed noise track (e.g., on an oscilloscope screen, Fig. 4.17) is not in excess of (5 to 6) σ_x.

This result, naturally, is close to the data given in Sec. 4.2 for the noise track in the case of a wide-band normal process (with the spectrum adjoining the zero frequency).

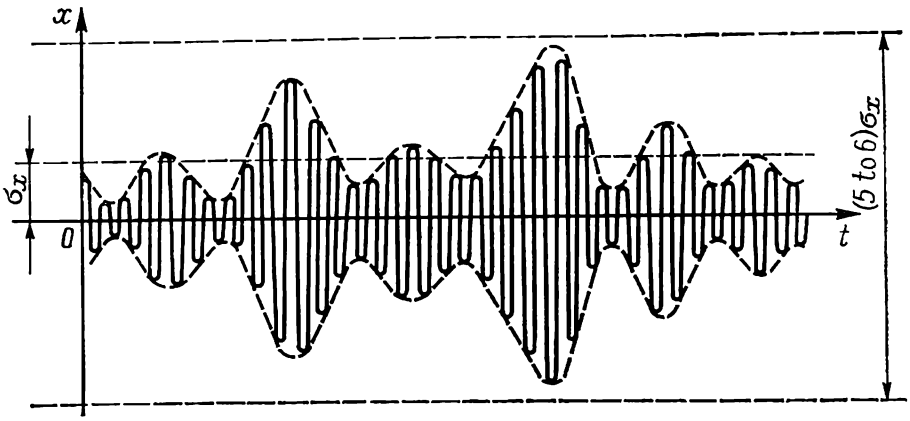


Fig. 4.17. Width of the noise track for a narrow-band noise with about 1-percent probability of exceeding the limits

The autocorrelation function of the envelope of narrow-band normal noise [6] is defined by the formula

$$B_A(\tau) = \frac{\pi\sigma_x^2}{2}\left\{1 + \left(\frac{1}{2}\right)^2 R_0^2(\tau) + \sum_{n=2}^{\infty}\left[\frac{1\times 3\ldots(2n-3)}{2\times 4\ldots 2n}\right]^2 R_0^{2n}(\tau)\right\} \tag{4.75}$$

Here $R_0(\tau)$ is the envelope of the normalized autocorrelation function of noise $x(t)$, i.e., the envelope of the function defined by the following expression (at $\bar{x} = 0$):

$$R_x(\tau) = B_x(\tau)/\sigma_x^2 = R_0(\tau)\cos\omega_0\tau \tag{4.76}$$

Since $R_0(\tau) \leqslant 1$, series (4.75) rapidly converges. Therefore, we may restrict ourselves to the first two terms:

$$B_A(\tau) \approx \frac{\pi\sigma_x^2}{2}\left[1 + \frac{1}{4}R_0^2(\tau)\right] \tag{4.77}$$

Applying the Fourier transformation to $B_A(\tau)$ [see (4.38)], we find the energy spectrum of the envelope:

$$W_A(\Omega) = \frac{\pi\sigma_x^2}{2}2\pi\delta(\Omega) + \frac{\pi\sigma_x^2}{2}\frac{1}{4}\int_{-\infty}^{\infty} R_0^2(\tau)\,e^{-i\Omega\tau}\,d\tau \tag{4.78}$$

From expression (4.78) it is clear that the energy spectrum of the envelope adjoins the zero frequency. The first term on the right-hand

side of (4.78) corresponds to the d-c component of the envelope, while the second one corresponds to the continuous part of the spectrum.

The examples of application of formulas (4.75) through (4.78) are given in Sec. 11.3 through 11.5.

4.6-2. Phase

Integrating the bivariate probability density $p(A, \theta)$, defined by expression (4.69) with respect to A gives the univariate probability density of phase

$$p_\theta(\theta) = \frac{1}{2\pi\sigma_x^2}\int_0^\infty A \exp\left(-\frac{A^2}{2\sigma_x^2}\right) dA$$

$$= \frac{1}{2\pi\sigma_x^2}\int_0^\infty \frac{1}{2} \exp\left(-\frac{A^2}{2\sigma_x^2}\right) d(A^2) = \frac{1}{2\pi}, \quad -\pi < \theta \leqslant \pi \qquad (4.79)$$

This result is in agreement with the integration limits in (4.70).

It should be noted that from the representation of $p(A, \theta)$ [see (4.69)] in the form of a product

$$p(A, \theta) = \frac{A}{2\pi\sigma_x^2} \exp\left(-\frac{A^2}{2\sigma_x^2}\right)$$

$$= \left[\frac{A}{\sigma_x^2} \exp\left(-\frac{A^2}{2\sigma_x^2}\right)\right]\left(\frac{1}{2\pi}\right) = p_A(A)\, p_\theta(\theta)$$

it directly follows that the random variables A and θ are statistically independent. As in the case of $A_c(t)$ and $A_s(t)$, this is valid when taking $A(t)$ and $\theta(t)$ at the same instant of time [see the remarks to (4.67)].

Relations (4.70) and (4.79) make it possible to draw the following general conclusion: a product of the type $x = A\cos\theta$, of which A and θ are statistically independent random variables, A obeying the Rayleigh distribution and θ being equiprobable in the range $(-\pi, \pi)$, has normal probability density.

The process $x(t)$ is not necessarily narrow-band; it is only necessary that A and θ should satisfy relations (4.63).

The autocorrelation function of the phase $\theta(t)$ is defined by the following expression [6]:

$$B_\theta(\tau) = \frac{\pi}{2} R_0(\tau) + \frac{1}{4} R_0^2(\tau) + \frac{\pi}{12} R_0^3(\tau) + \ldots \qquad (4.80)$$

At $\tau = 0$ the series converges to $\pi^2/3$, i.e., the phase variance is equal to $\pi^2/3$. Indeed, for distribution (4.79),

$$\sigma_\theta^2 = \int_{-\pi}^{\pi} \theta^2 p_\theta(\theta)\, d\theta = \frac{1}{2\pi}\left(\frac{\theta^3}{3}\right)_{-\pi}^{\pi} = \frac{\pi^2}{3} \tag{4.81}$$

4.6-3. Frequency

On the basis of (4.60), the instantaneous frequency of noise may be written in the form

$$\omega(t) = \frac{d\psi(t)}{dt} = \omega_0 + \frac{d\theta(t)}{dt} = \omega_0 + \dot{\theta}(t)$$

from which it is clear that the distribution of the instantaneous frequency is defined by the distribution of the derivative of the phase θ.

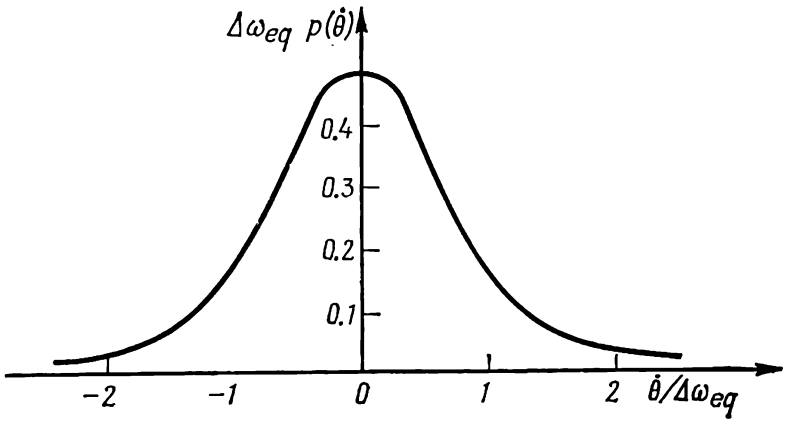


Fig. 4.18. Probability density of the derivative of the phase of a normal random process

Let us give (without proof) an expression for the probability density of the random variable $\dot{\theta}$ [6, 7]:

$$p(\dot{\theta}) = \frac{1}{2}\Big/\Delta\omega_{eq}\left(1 + \frac{\dot{\theta}^2}{\Delta\omega_{eq}^2}\right)^{3/2} \tag{4.82}$$

where $\Delta\omega_{eq}$ is the equivalent spectrum width of a narrow-band process, defined by the expression

$$\Delta\omega_{eq}^2 = \int_0^\infty (\omega - \omega_0)^2 W(\omega)\, d\omega \Big/ \int_0^\infty W(\omega)\, d\omega \tag{4.83}$$

The last expression is equivalent to the formula

$$\Delta\omega_{eq}^2 = -\frac{d^2 R_0(\tau)}{d\tau^2}\Big|_{\tau=0} \tag{4.83'}$$

where $R_0(\tau)$ is the envelope of the normalized autocorrelation function of a process having an energy spectrum $W(\omega)$ [symmetrical with respect to the centre frequency ω_0].

The graph of the function $p(\dot{\theta})$ is shown in Fig. 4.18. The mean value of the magnitude $|\dot{\theta}|$ is equal to $\Delta\omega_{eq}$.

Let us consider an exemplary case where the energy spectrum $W(\omega)$ is uniform within the frequency band $\pm \Delta\omega$ with a centre frequency ω_0.

The normalized autocorrelation function, in accordance with (4.44), is

$$R_x(\tau) = \frac{\sin \Delta\omega_0 \tau}{\Delta\omega_0 \tau} \cos \omega_0 \tau$$

and

$$R_0(\tau) = \frac{\sin \Delta\omega_0 \tau}{\Delta\omega_0 \tau} = \frac{\sin y}{y}$$

Differenting doubly the last expression with respect to τ, we obtain

$$R_0''(\tau) = \Delta\omega_0^2 \frac{-y^3 \sin y - 2y^2 \cos y + 2y \sin y}{y^4}$$

When $\tau \to 0$ and $y \to 0$, we have

$$R_0''(0) = -\frac{1}{3} \Delta\omega_0^2$$

and

$$\Delta\omega_{eq} = \sqrt{-R_0''(0)} = \Delta\omega_0 / \sqrt{3} \qquad (4.85'')$$

Thus, in the case of noise whose energy spectrum is uniformly distributed in the band $(-\Delta\omega_0, \Delta\omega_0)$ [see Fig. 4.9], the mean value of the magnitude $|\dot{\theta}|$ is equal to $\Delta\omega_0/\sqrt{3}$.

4.7. OSCILLATION MODULATED IN AMPLITUDE BY A RANDOM PROCESS

Let us return to expression (3.4) and consider it from probabilistic point of view, taking into account that the message being transmitted is contained in the envelope $A(t)$.

Let the envelope $A(t)$ be a stationary ergodic random process. The relation between the change in the envelope and the modulating function $s(t)$ is defined, as in Sec. 3.2, by the expression $\Delta A(t) = k_{am} s(t)$, where k_{am} has the meaning of the slope of characteristic curve of the amplitude modulator.

The distribution law of the message to be transmitted is taken to be normal with zero mean. Thus, the probability density of the variable $s(t)$ is

$$p(s) = \frac{1}{\sqrt{2\pi}\,\sigma_s} \exp\left(-\frac{s^2}{2\sigma_s^2}\right) \tag{4.84}$$

Respectively, the probability density of the envelope A in *linear modulation* can be taken to be normal:

$$p(A) = \frac{1}{\sqrt{2\pi}\,\sigma_A} \exp\left[-\frac{(A-A_0)^2}{2\sigma_A^2}\right] \tag{4.85}$$

where $\sigma_A = k_{am}\sigma_s$.

Expression (4.85) holds and makes sense if the root-mean-square value of the modulation factor is not high enough for the limitation

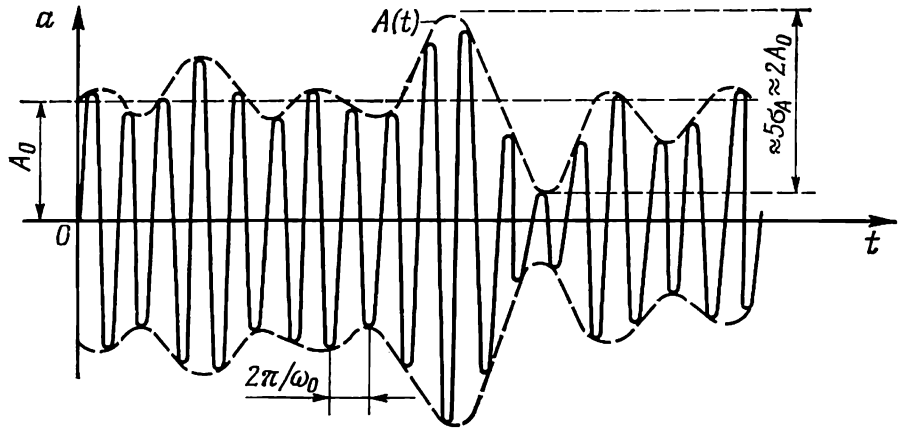


Fig. 4.19. General view of a radio-frequency oscillation modulated in amplitude by a normal random process

of the envelope in downward modulation to become evident. Under the root-mean-square (rms) modulation factor we may imply the ratio $M_{rms} = \sigma_A/A_0^*$. It is required that the probability that $A(t)$ falls down to zero should be negligibly low. This requirement is met if $M_{rms} \leqslant 0.3$ to 0.4 (see Sec. 4.2, Para. 4), i.e., if the width of the noise track (over the envelope), approximately equal to $5\sigma_A$, is not in excess of twice the amplitude A_0. A general view of one of the realizations of the modulated wave $a(t)$ satisfying the above condition is shown in Fig. 4.19.

---

* It should be noted that in harmonic modulation according to $A(t) = A_0(1 + M\cos\Omega t)$, the root-mean-square value M_{rms} is determined from the relation $M_{rms} = \overline{(M\cos\Omega t)^2} = M^2/2$, whence $M_{rms} = M/\sqrt{2}$.

Let us select an arbitrary moment of time $t = t_1$ and write expression (3.4) for this moment:

$$a(t_1) = A(t_1) \cos(\omega_0 t_1 + \theta_0) \tag{4.86}$$

Multiplying the random variable $A(t_1)$, whose univariate probability density $p(A)$ is independent of the moment of time selected, by the deterministic factor $\cos(\omega_0 t_1 + \theta_0)$ results in a change in the mathematical expectation and variance, depending on the choice of t_1. The random process $a(t)$ defined by expression (4.86) remains normal, but it is not stationary and consequently, not ergodic (see Sec. 4.2, Para. 2, where we discussed a similar random process, but with other envelope distribution).

For complete description of the random process $a(t)$, let us find its autocorrelation function and energy spectrum. Since the process is not stationary, its autocorrelation function depends both on the interval τ and on the time t:

$$B_a(t, \tau) = \langle a(t)\, a(t+\tau)\rangle = \langle A(t)\, A(t+\tau)\rangle \cos(\omega_0 t + \theta_0) \cos(\omega_0 t + \omega_0 \tau + \theta_0)$$
$$= \langle A(t)\, A(t+\tau)\rangle \frac{1}{2} [\cos \omega_0 \tau + \cos(2\omega_0 t + 2\theta_0 + \omega_0 \tau)] \tag{4.87}$$

Since $A(t)$ is a stationary process, the term

$$\langle A(t)\, A(t+\tau)\rangle = B_A(\tau) \tag{4.88}$$

in (4.87) is the autocorrelation function of the envelope $A(\tau)$, which is independent of t.

The quantity $\cos(2\omega_0 t + 2\theta_0 + \omega_0 \tau)$ that depends on t is a rapidly oscillating (with a frequency $2\omega_0$) function. In such cases use is made of the *time average* of the autocorrelation function

$$B_a(\tau) = \overline{B_a(t, \tau)} = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} B_a(t, \tau)\, dt \tag{4.89}$$

With such an averaging, the term $\cos(2\omega_0 t + 2\theta_0 + \omega_0 \tau)$ in (4.87) may be omitted. Thus, in the example under consideration

$$B_a(\tau) = \frac{1}{2} B_A(\tau) \cos \omega_0 \tau \tag{4.90}$$

This result coincides with expression (3.103) derived for a deterministic oscillation, the only difference being in the method of defining $B_A(\tau)$. In Sec. 3.11, $B_A(0)$ had the dimension of energy, while in the given case $B_A(0)$ has the meaning of the average power of the random process.

Applying expression (4.38) to the time average of the function $B_a(\tau)$, we find the energy spectrum of the modulated wave:

$$W_a(\omega) = \int_{-\infty}^{\infty} B_a(\tau)\, e^{-i\omega\tau}\, d\tau = \frac{1}{2} \int_{-\infty}^{\infty} B_A(\tau) \cos\omega_0\tau e^{-i\omega\tau}\, d\tau$$

$$= \frac{1}{4} \int_{-\infty}^{\infty} B_A(\tau)\, e^{-i(\omega-\omega_0)\tau} d\tau + \frac{1}{4} \int_{-\infty}^{\infty} B_A(\tau)\, e^{-i(\omega+\omega_0)\tau}\, d\tau \qquad (4.91)$$

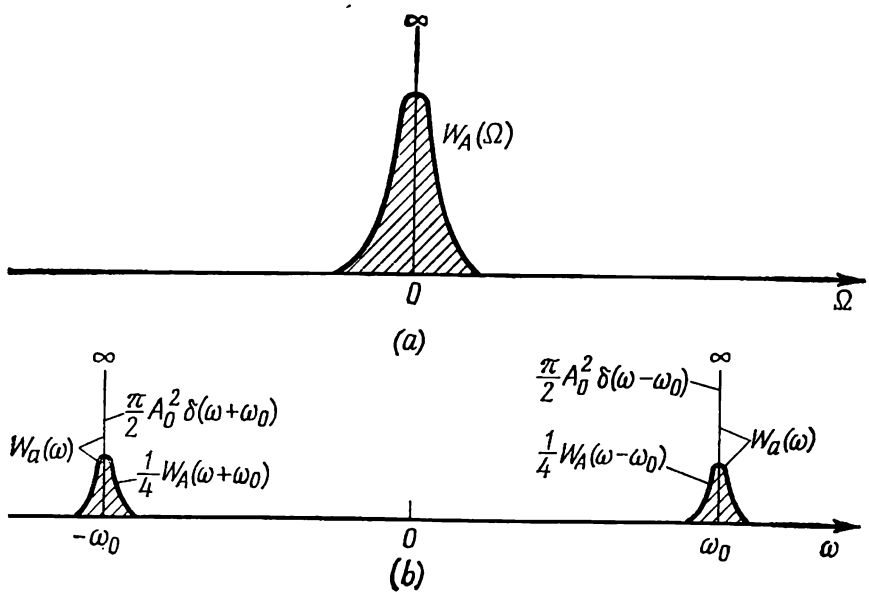


Fig. 4.20. Energy spectra of (*a*) the envelope in amplitude modulation by a normal random process and (*b*) the instantaneous value of the modulated oscillation

Taking into account that the mean value of $A(t)$ is equal to A_0, let us represent $B_A(\tau)$ in the form of expression (4.18), replacing therein $\bar{x}$ by A_0, $\rho_x(\tau)$ by $\rho_A(\tau)$, and $B_x(\tau)$ by $B_A(\tau)$:

$$B_A(\tau) = A_0^2 + \rho_A(\tau) \qquad (4.92)$$

where $\rho_A(\tau)$ is the autocovariance function of the envelope.

Substituting (4.92) into (4.91) and taking into account (2.94′), we obtain

$$W_a(\omega) = \frac{A_0^2}{4} 2\pi [\delta(\omega-\omega_0) + \delta(\omega+\omega_0)]$$

$$+ \frac{1}{4} [W_A(\omega-\omega_0) + W_A(\omega+\omega_0)] \qquad (4.93)$$

In this expression, $W_A(\omega - \omega_0) = W_A(\Omega)$ is the energy spectrum of the fluctuating part of the envelope $A(t)$, related to the energy spectrum $W_s(\Omega)$ of the modulating function $s(t)$ by the obvious relation $W_A(\Omega) = k_{am}^2 W_s(\Omega)$. The relation between $W_A(\Omega)$ and $W_a(\omega)$ is illustrated in Fig. 4.20.

4.8. OSCILLATION MODULATED IN PHASE BY A RANDOM PROCESS. PROBABILITY DENSITY

Let a modulated wave be represented in the form

$$a(t) = A_0 \cos [w_0 t + \theta(t) + \theta_0] \tag{4.94}$$

As in the preceding section, let the modulating function $s(t)$ have a normal distribution [see (4.84)].

With the phase modulator having a linear characteristic of slope k_{pm} [rad/V], the instantaneous phase is defined by the expression

$$\theta(t) = k_{pm} s(t) \tag{4.95}$$

and the probability density of the random variable θ is expressed as

$$p(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma_\theta} \exp\left(-\frac{\theta^2}{2\sigma_\theta^2}\right) \tag{4.96}$$

where

$$\sigma_\theta = \sqrt{\overline{\theta^2(t)}} = m_{rms} = k_{pm}\sigma_s \tag{4.97}$$

may be regarded as the root-mean-square value of the modulation index.

It should be noted that in harmonic phase modulation according to $\theta(t) = m \sin \Omega t$, obviously,

$$\overline{\theta^2(t)} = m^2/2, \text{ i. e., } m_{rms} \leqslant m/\sqrt{2} \tag{4.98}$$

When determining the probability density of the function $a(t)$, we should distinguish between two cases: (a) deterministic epoch angle θ_0 and (b) random epoch angle θ_0.

Let us consider the first case. When measuring a voltage $a(t)$ at any fixed instant t_1, its value $a(t_1)$ differs from the deterministic value $A_0 \cos(\omega_0 t_1 + \theta_0)$ only due to the presence of a random phase shift $\theta(t)$ caused by the modulation. Setting the probability density of $\theta(t)$ makes it possible to find the probability density of $a(t)$ by the same reasoning as was used when deriving formula (4.25), the only difference being that here the value of $\theta(t)$ is not limited in the interval $(-\pi, \pi)$ and, besides, is not equiprobable in any interval.

When calculating the probability of residence of $a(t)$ in a given interval $(a, a + da)$, one should take into account all the phase intervals in which the probability density $p(\theta)$ differs from zero. For example, if the epoch angle $\theta_0 = 0$, the disposition of these

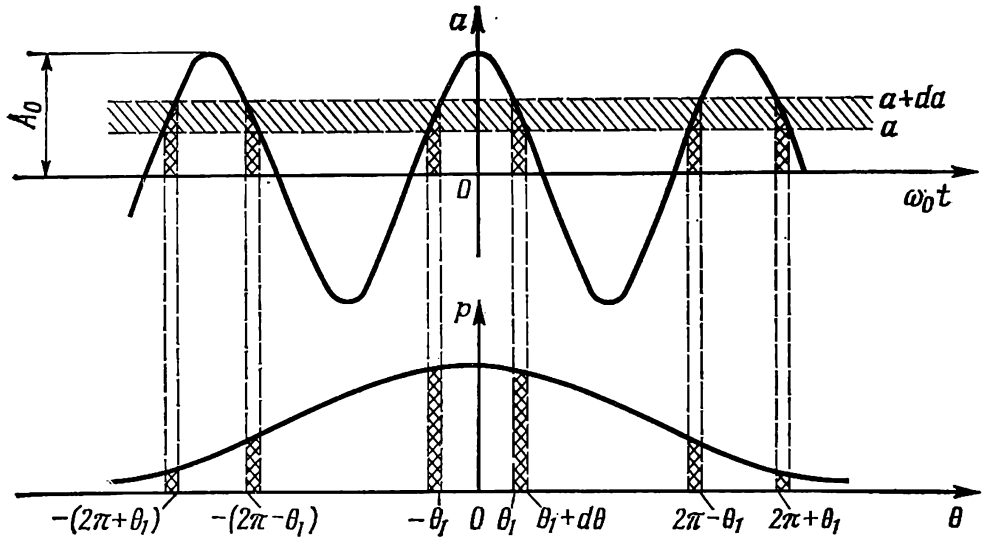


Fig. 4.21. Determining the probability density of a radio-frequency oscillation modulated in phase by a random process

intervals corresponds to that shown in Fig. 4.21. In this particular case the probability density $p(a)$ [by analogy with (4.25) and with account being taken of (4.96)] assumes the following form

$$p(a) = \sum_i \frac{p(\theta)_i}{\left|\frac{da}{d\theta}\right|_i} = \frac{\frac{2}{\sqrt{2\pi}\,\sigma_\theta}\sum_i \exp\left(-\frac{\theta_i^2}{2\sigma_\theta^2}\right)}{\sqrt{A_0^2 - a^2}}$$

$$= \frac{2}{\sqrt{2\pi}\,\sigma_\theta}\,\frac{1}{\sqrt{A_0^2 - a^2}}\left\{\exp\left(-\frac{\theta_1^2}{2\sigma_\theta^2}\right) + \exp\left[-\frac{(2\pi-\theta_1)^2}{2\sigma_\theta^2}\right]\right.$$

$$\left. + \exp\left[-\frac{(2\pi+\theta_1)^2}{2\sigma_\theta^2}\right] + \exp\left[-\frac{(4\pi-\theta_1)^2}{2\sigma_\theta^2}\right] + \ldots\right\}, \quad -A_0 < a < A_0 \tag{4.99}$$

In this expression, $\theta_1 = \arccos(a/A_0)$, $0 \leqslant |\theta_1| \leqslant \pi$. The number of significant terms in (4.99) depends on the root-mean-square value σ_θ of the modulation index. At relatively low σ_θ (1 to 2 rad), we may restrict ourselves to the first term. The graphs of $p(a)$ for several σ_θ are shown in Fig. 4.22.

It is clear that as $\sigma_\theta \to 0$, the distribution approaches $p(a) \to \delta(a - A_0)$ and this corresponds to a 100-percent probability that

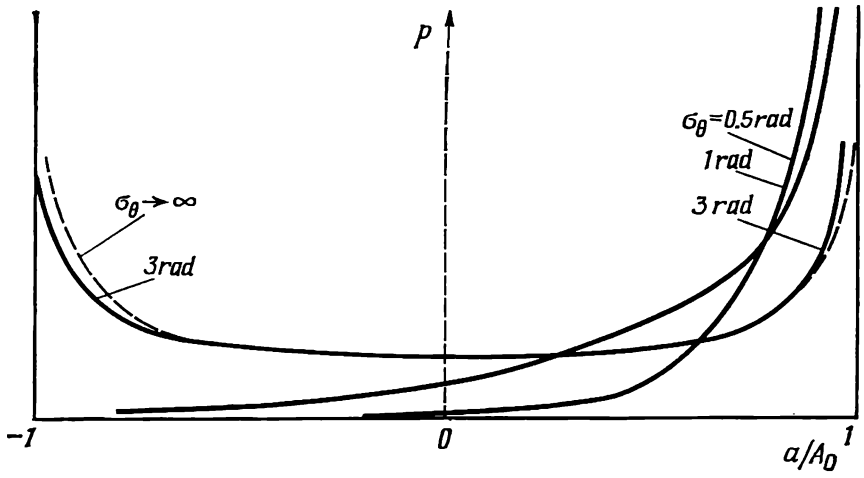


Fig. 4.22. Probability density of a radio-frequency oscillation modulated in phase by a random process

$a(t)$ has a peak value at the moment $t = 0$. At relatively high σ_θ (in excess of 3 rad), the distribution of the random variable $a(t)$ differs but slightly from that corresponding to a harmonic oscillation

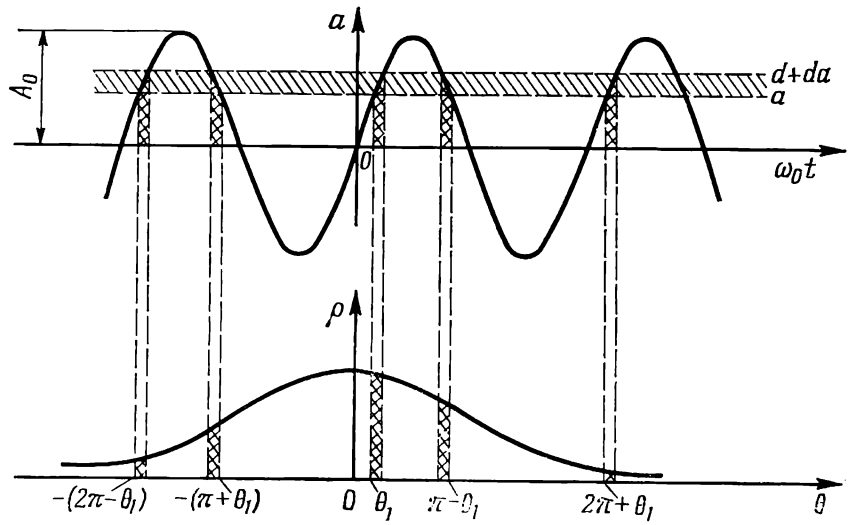


Fig. 4.23. The same as in Fig. 4.21 but with a different epoch angle of the carrier oscillation

whose phase is equiprobable in the interval $(0, 2\pi)$. The graph of $p(a)$ in this case almost coincides with the graph plotted by formula (4.25) (dashed line in Fig. 4.22).

The picture becomes essentially different when the epoch angle θ_0 of the unmodulated wave is equal to $\pi/2$. The relation between the intervals of values a (t), the associated phase intervals, and the probability density p (θ) is shown in Fig. 4.23. At $\theta_0 = \pi/2$, the probability density p (a) is defined by a formula similar to (4.99):

$$p(a) = \frac{1}{\sqrt{2\pi}\,\sigma_\theta}\frac{1}{\sqrt{A_0^2 - a^2}}\left\{\exp\left(-\frac{\theta_1^2}{2\sigma_\theta^2}\right)\right.$$
$$+\exp\left[-\frac{(\pi-\theta_1)^2}{2\sigma_\theta^2}\right]+\exp\left[-\frac{(\pi+\theta_1)^2}{2\sigma_\theta^2}\right]+\exp\left[-\frac{(2\pi-\theta_1)^2}{2\sigma_\theta^2}\right]$$
$$\left.+\exp\left[-\frac{(2\pi+\theta_1)^2}{2\sigma_\theta^2}\right]+\ldots\right\},\quad \theta_1 = \arcsin(a/A_0),\ |\theta| \leqslant \pi/2 \quad (4.100)$$

The graphs of p (a) for several σ_θ are shown in Fig. 4.24.

At low σ_θ, the distribution of a (t) approaches normal, and at high σ_θ, it approaches distribution (4.25).

From comparison between the two typical conditions, $\theta_0 = 0$ and $\theta_0 = \pi/2$, it is clear that the probability density p (a) depends on the epoch angle or, which is the same, on the reference time t_1. Thus, with a deterministic epoch angle θ_0, $a(t)$ is a *nonstationary process.* However, it should be noted that the influence of θ_0 on p (a) weakens with an increase in σ_θ. At sufficiently high σ_θ, the process approaches a stationary one.

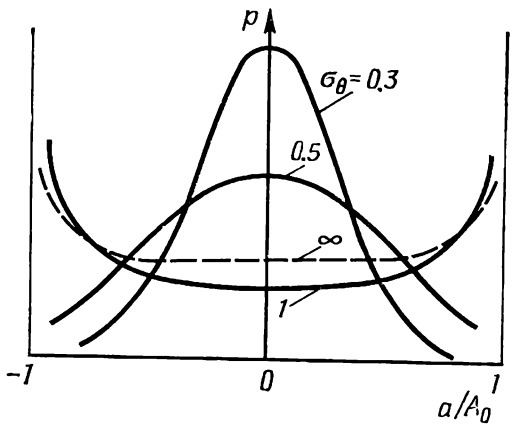


Fig. 4.24. The same as in Fig. 4.22 but with a different initial phase of the carrier oscillation

The distribution of a (t) in phase modulation with a random epoch angle θ_0 is not discussed here. It should only be noted that in sufficiently slow modulation satisfying condition (3.3), i.e., when the oscillation a (t) preserves a shape close to a harmonic oscillation, the probability density p (a) coincides with (4.25).

We do not consider here the autocorrelation function and the energy spectrum in angle modulation by a random process (see [6]).

Chapter 5

LINEAR RADIO CIRCUITS WITH CONSTANT PARAMETERS

5.1. GENERAL

This chapter gives fundamentals of linear active circuits. Considered here are the frequency characteristics of selective circuits used for various linear conversion operations (amplification, filtering, etc.) on signals. Special attention is given to the study of linear active circuits with feedback, that most of the modern radioelectronic devices make use of.

The presentation of the material is based on the fundamentals of the electric circuit theory that the students must already be acquainted with.

5.2. FUNDAMENTAL PROPERTIES OF ACTIVE NETWORKS. DEFINITIONS

In the general electric circuit theory, the term active circuit implies a network including not only passive elements (inductors, capacitors, and resistors) but also power sources (voltage or current generators).

The active character of radioelectronic circuits derives from the fact that they use amplifying elements — electron tubes, transistors, travelling-wave tubes, etc.— it being presumed that the power level of a signal at the output of an active network is higher than that at its input. To make the definition more strict, let us change it as follows: *a circuit is active if, in harmonic excitation, the average signal power at the output of the circuit is higher than that at its input*, i.e., if the power gain is higher than unity. From this definition it is clear that a circuit which amplifies voltage without amplifying power, e.g., by means of a step-up transformer, is passive, even though it may include active elements with power sources of their own.

When constructing the equivalent circuits of active networks, power sources are omitted. In these circuits, the active elements (tubes, transistors, etc.) are shown by means of equivalent parameters which depend on the operating conditions of the elements and, in the final analysis, on the power sources supplying them. Under these assumptions, any (active or passive) linear two-port network

can be shown schematically as in Fig. 5.1. In this figure, $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{I}_1$ and $\mathbf{I}_2$ stand for the complex amplitudes of harmonic voltages and currents of independent sources at a fixed frequency ω.

A two-port network is characterized by the relations between the voltages and currents at its input and output. The form of these relations depends on the choice of the independent variables.

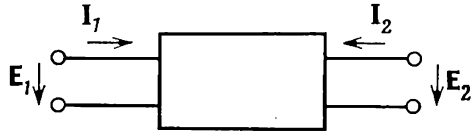


Fig. 5.1. Equivalent circuit of a linear two-port network

Let us briefly recall the basic forms of representation of two-port networks.

If $\mathbf{E}_1$ and $\mathbf{E}_2$ are independent variables, the equations for the currents $\mathbf{I}_1$ and $\mathbf{I}_2$ are written in the form

$$\left.\begin{aligned}\mathbf{I}_1 &= Y_{11}\mathbf{E}_1 + Y_{12}\mathbf{E}_2\\ \mathbf{I}_2 &= Y_{21}\mathbf{E}_1 + Y_{22}\mathbf{E}_2\end{aligned}\right\} \tag{5.1}$$

or, in matrix notation,

$$\begin{bmatrix}\mathbf{I}_1\\ \mathbf{I}_2\end{bmatrix} = [Y]\begin{bmatrix}\mathbf{E}_1\\ \mathbf{E}_2\end{bmatrix} \tag{5.2}$$

where

$$[Y] = \begin{bmatrix}Y_{11} & Y_{12}\\ Y_{21} & Y_{22}\end{bmatrix} \tag{5.3}$$

is a matrix of parameters having the meaning and dimensions of admittance.

If equations (5.1) are solved for $\mathbf{E}_1$ and $\mathbf{E}_2$ in terms of $\mathbf{I}_1$ and $\mathbf{I}_2$, one gets the following set of equations:

$$\left.\begin{aligned}\mathbf{E}_1 &= Z_{11}\mathbf{I}_1 + Z_{12}\mathbf{I}_2\\ \mathbf{E}_2 &= Z_{21}\mathbf{I}_1 + Z_{22}\mathbf{I}_2\end{aligned}\right\} \tag{5.4}$$

$$\begin{bmatrix}\mathbf{E}_1\\ \mathbf{E}_2\end{bmatrix} = [Z]\begin{bmatrix}\mathbf{I}_1\\ \mathbf{I}_2\end{bmatrix} \tag{5.5}$$

where

$$[Z] = \begin{bmatrix}Z_{11} & Z_{12}\\ Z_{21} & Z_{22}\end{bmatrix} \tag{5.6}$$

is a matrix of parameters having the dimensions of impedance.

The two-port equations written in the form

$$\left.\begin{aligned} \mathbf{E}_1 &= H_{11}\mathbf{I}_1 + H_{12}\mathbf{E}_2 \\ \mathbf{I}_2 &= H_{21}\mathbf{I}_1 + H_{22}\mathbf{E}_2 \end{aligned}\right\} \tag{5.7}$$

have a corresponding matrix of parameters

$$[H] = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{5.8}$$

in which H_{11} has the dimension of impedance and H_{22}, that of admittance, while H_{12} and H_{21} are dimensionless parameters.

Let us write the two-port equations in still another form

$$\left.\begin{aligned} \mathbf{I}_1 &= G_{11}\mathbf{E}_1 + G_{12}\mathbf{I}_2 \\ \mathbf{E}_2 &= G_{21}\mathbf{E}_1 + G_{22}\mathbf{I}_2 \end{aligned}\right\} \tag{5.9}$$

with a corresponding matrix

$$[G] = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \tag{5.10}$$

where G_{11} has the dimension of admittance and G_{22} that of impedance, while G_{12} and G_{21} are dimensionless parameters.

The matrices of the Z-, Y- and H-parameters (called respectively the impedance, admittance, and hybrid matrices) are most widely used in the theory of amplifiers. The interrelations between the above

Table 5.1

Admittance Parameter Set	Relation to Other Parameter Sets	Impedance Parameter Set	Relation to Other Parameter Sets	Hybrid Parameter Set	Relation to Other Parameter Sets
Y_{11}	$\frac{Z_{22}}{\Delta Z}\quad\frac{1}{H_{11}}$	Z_{11}	$\frac{Y_{22}}{\Delta Y}\quad\frac{\Delta H}{H_{22}}$	H_{11}	$\frac{\Delta Z}{Z_{22}}\quad\frac{1}{Y_{11}}$
Y_{12}	$-\frac{Z_{12}}{\Delta Z}\quad-\frac{H_{12}}{H_{11}}$	Z_{12}	$-\frac{Y_{12}}{\Delta Y}\quad\frac{H_{12}}{H_{22}}$	H_{12}	$\frac{Z_{12}}{Z_{22}}\quad-\frac{Y_{12}}{Y_{11}}$
Y_{21}	$-\frac{Z_{21}}{\Delta Z}\quad\frac{H_{21}}{H_{11}}$	Z_{21}	$-\frac{Y_{21}}{\Delta Y}\quad-\frac{H_{21}}{H_{22}}$	H_{21}	$-\frac{Z_{21}}{Z_{22}}\quad\frac{Y_{21}}{Y_{11}}$
Y_{22}	$\frac{Z_{11}}{\Delta Z}\quad\frac{\Delta H}{H_{11}}$	Z_{22}	$\frac{Y_{11}}{\Delta Y}\quad\frac{1}{H_{22}}$	H_{22}	$\frac{1}{Z_{22}}\quad\frac{\Delta Y}{Y_{11}}$

parameter sets are given in Table 5.1. In this table, the determinants, ΔY, ΔZ, and ΔH, of the respective matrices are defined by the

expressions

$$\left.\begin{aligned}\Delta Y &= Y_{11}Y_{22} - Y_{12}Y_{21}\\ \Delta Z &= Z_{11}Z_{22} - Z_{12}Z_{21}\\ \Delta H &= H_{11}H_{22} - H_{12}H_{21}\end{aligned}\right\} \quad (5.11)$$

Equations (5.1), (5.4), and (5.7), and also other similar equations make it possible to construct equivalent circuits for various two-port networks.

Figure 5.2*a* shows an equivalent circuit* constructed in accordance with equation (5.1). In this circuit, both voltages, $\mathbf{E}_1$ and $\mathbf{E}_2$, are regarded as voltages from external sources. The current generator $Y_{12}\mathbf{E}_2$ takes into account the effect of the output voltage $\mathbf{E}_2$ on the input current $\mathbf{I}_1$, while the current generator $Y_{21}\mathbf{E}_1$ takes account of the effect of the input voltage $\mathbf{E}_1$ on the output current $\mathbf{I}_2$. Both generators may be regarded as "dependent sources", since the currents provided by them are proportional to the voltages of the external sources. The parameter Y_{21} is termed the forward-transfer admittance, i.e., from the input to the output, while Y_{12}, the reverse-transfer admittance, i.e., from the output to the input. It is obvious that Y_{11} is the input admittance of the two-port network at $\mathbf{E}_2 = 0$, i.e., with the output terminals short circuited, and Y_{22} is the input admittance of the two-port network at $\mathbf{E}_1 = 0$, i.e., with the input terminals short circuited.

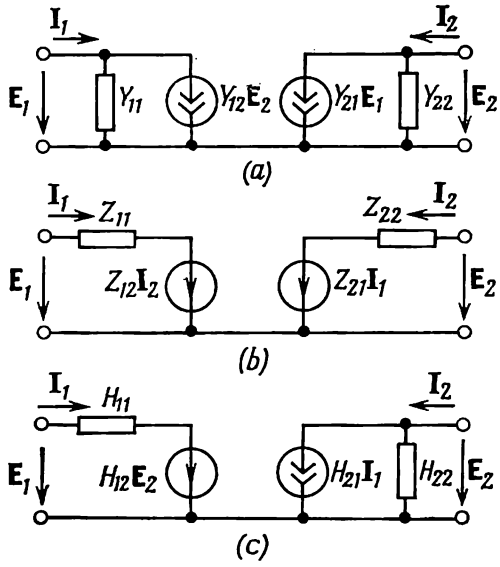


Fig. 5.2. Equivalent circuits of a two-port network, based on (a) *Y*-parameter matrix, (*b*) Z-parameter matrix and (*c*) *H*-parameter matrix

The equivalent circuit of the two-port network corresponding to equations (5.4) and (5.5) is shown in Fig. 5.2*b*. In this diagram, the dependent voltage sources $Z_{12}\mathbf{I}_2$ and $Z_{21}\mathbf{I}_1$ take into account the influence of $\mathbf{I}_2$ on $\mathbf{E}_1$ and of $\mathbf{I}_1$ on $\mathbf{E}_2$, respectively. The equiva-

* The presence of a common conductor in the circuit of Fig. 5.2 and in similar circuits allows one to speak of a thee-terminal network. This does not affect the network equations.

lent circuit shown in Fig. 5.2*c* corresponds to equations (5.7) and (5.8).

It should be noted here that active two-port networks possess the following specific feature: as a rule, $Y_{21} \neq Y_{12}$ or $Z_{21} \neq Z_{12}$, and $H_{21} \neq H_{12}$. This means that active two-port networks are *nonreciprocal* and therefore, the reciprocity principle cannot be applied to active two-ports.

It is known that the transfer admittances or impedances of passive two-port networks are equal (the reciprocity theorem). This allows equivalent circuits, e.g., those shown in Fig. 5.2*a* and *b*,

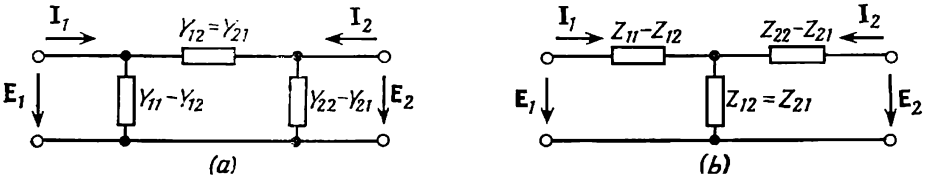


Fig. 5.3. Conversion of the equivalent circuits shown in Fig. 5.2*a* and which holds only for a passive two-port network

to be simplified and reduced to forms containing no dependent sources (Fig. 5.3), provided the two-port is passive.

When analyzing radio circuits, one often has to deal with two-port networks excited at the input only; by the output voltage in this case is meant the voltage drop across the load impedance $Z_l = 1/Y_l$, i.e., $\mathbf{E}_2 = -\mathbf{I}_2 Y_l$. In such cases the load element is preferably inserted inside the two-port.

If a two-port network is represented in terms of the Y-matrix, we have an equivalent circuit shown in Fig. 5.4*a*, which differs from the circuit shown in Fig. 5.2*a* only in that the load admittance Y_l is added to Y_{22}. This makes it possible to consider the new two-port as an open-circuit network having an output current $\mathbf{I}_2' = 0$. The parameter matrix of this new two-port network is

$$[Y]' = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22}' \end{bmatrix} \tag{5.12}$$

where $Y_{22}' = Y_{22} + Y_l$.

The second equation of (5.1) in this case assumes the form

$$\mathbf{I}_2' = Y_{21}\mathbf{E}_1 + Y_{22}'\mathbf{E}_2 = 0$$

from which follows an important relation

$$\frac{\mathbf{E}_2}{\mathbf{E}_1} = -\frac{Y_{21}}{Y_{22}'} = -\frac{Y_{21}}{Y_{22}+Y_l} \tag{5.13}$$

Using this relation to exclude $\mathbf{E}_1$ from the first equation of (5.1) and taking into account that $\mathbf{E}_2 = \mathbf{I}_2 Z_l$, we obtain the following current ratio

$$\frac{\mathbf{I}_2}{\mathbf{I}_1} = \frac{Y_{21} Y_l}{Y_{11} Y'_{22} - Y_{12} Y_{21}} = \frac{Y_{21} Y_l}{\Delta Y'} \tag{5.13'}$$

where $\Delta Y' = Y_{11} (Y_{22} + Y_l) - Y_{12} Y_{21}$ is the determinant of matrix (5.12).

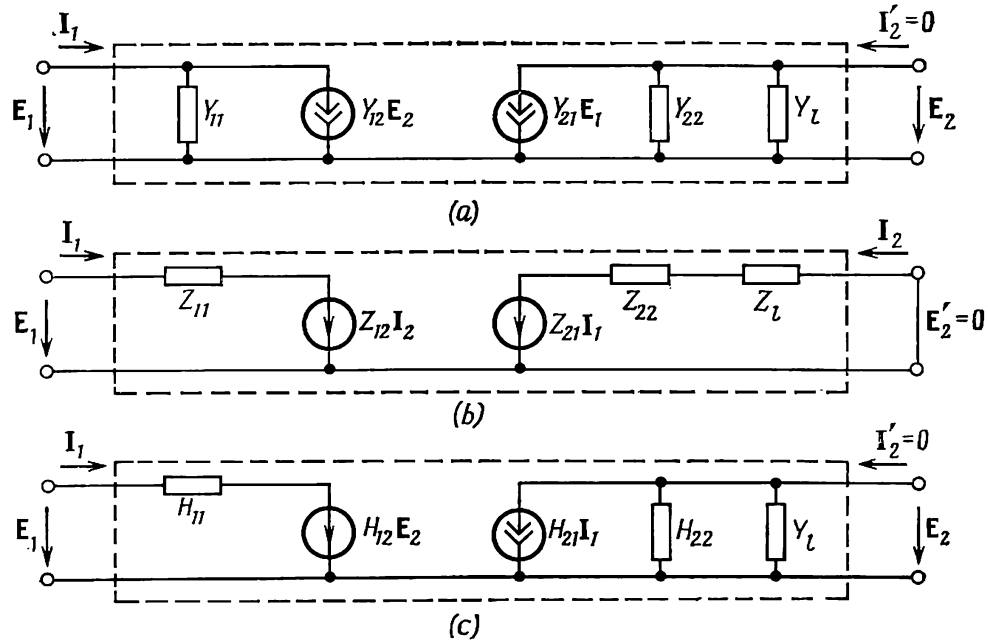


Fig. 5.4. Incorporating the load element into a two-port network

When using the Z-matrix, one has an equivalent circuit shown in Fig. 5.4*b*. In this case the output terminals are short circuited ($\mathbf{E}'_2 = 0$), and the parameter matrix becomes

$$[Z]' = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12} & Z'_{22} \end{bmatrix}$$

where $Z'_{22} = Z_{22} + Z_l$.

The second equation of (5.4) in this case yields the relation

$$\frac{\mathbf{I}_2}{\mathbf{I}_1} = -\frac{Z_{21}}{Z'_{22}} = -\frac{Z_{21}}{Z_{22} + Z_l} \tag{5.14}$$

and the first equation, the relation

$$\frac{\mathbf{E}_2}{\mathbf{E}_1} = \frac{Z_{21} Z_l}{Z_{11} Z'_{22} - Z_{12} Z_{21}} = \frac{Z_{21} Z_l}{\Delta Z'} \tag{5.14'}$$

where $\Delta Z' = Z_{11}(Z_{22} + Z_l) - Z_{12}Z_{21}$ is the determinant of the matrix $[Z]'$.

Finally, the second equation of (5.7), after substituting $H'_{22} = H_{22} + Y_l$ and $\mathbf{E}_2 = -\mathbf{I}_2 Z_l$ (Fig. 5.4*c*), gives

$$0 = H_{21}\mathbf{I}_1 + H'_{22}\mathbf{E}_2 = H_{21}\mathbf{I}_1 - H'_{22}Z_l\mathbf{I}_2$$

from which follows the relation

$$\frac{\mathbf{I}_2}{\mathbf{I}_1} = \frac{H_{21}}{H'_{22}Z_l} = \frac{H_{21}}{(H_{22}+Y_l)Z_l} = \frac{H_{21}}{Z_l H_{22}+1} \tag{5.15}$$

Using this relation to exclude $\mathbf{I}_1$ from the first equation of (5.7), we obtain

$$\frac{\mathbf{E}_2}{\mathbf{E}_1} = \frac{H_{21}}{H_{11}H'_{22} - H_{12}H_{21}} = \frac{H_{21}}{\Delta H'} \tag{5.15'}$$

where $\Delta H' = H_{11}(H_{22} + Y_l) - H_{12}H_{21}$.

The general equations (5.1), (5.4) and (5.7) can be transformed so that the respective equivalent circuits of the two-port network will include only one dependent source.

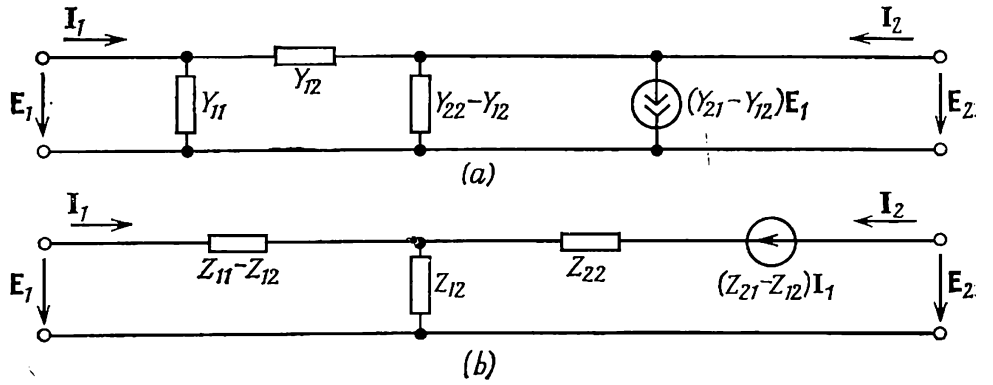


Fig. 5.5. Equivalent circuits with one dependent source of (*a*) current and (*b*) voltage

For example, by writing the second equation of (5.1) in the form

$$\mathbf{I}_2 = Y_{12}\mathbf{E}_1 + Y_{22}\mathbf{E}_2 + (Y_{21} - Y_{12})\mathbf{E}_1 \tag{5.16}$$

we come to an equivalent circuit (Fig. 5.5*a*) including a single dependent current source, $(Y_{21} - Y_{12})\mathbf{E}_1$.

In a similar way, by writing the second equation of (5.4) in the form

$$\mathbf{E}_2 = Z_{12}\mathbf{I}_1 + Z_{22}\mathbf{I}_2 + (Z_{21} - Z_{12})\mathbf{I}_1 \tag{5.16'}$$

we come to a circuit with a single dependent voltage source, $(Z_{21} - Z_{12})\mathbf{I}_1$ (Fig. 5.5*b*).

5.3. ACTIVE TWO-PORT NETWORK AS A LINEAR AMPLIFIER

Expressions (5.13) through (5.15′) of the preceding section, when written in the form

$$\mathrm{K}_E=\frac{\mathrm{E}_2}{\mathrm{E}_1}=\frac{-Y_{21}}{Y_{22}+Y_l}=\frac{Z_{21}Z_l}{Z_{11}(Z_{22}+Z_l)-Z_{12}Z_{21}}=-\frac{H_{21}}{H_{11}(H_{22}+Y_l)-H_{12}H_{21}} \tag{5.17}$$

$$\mathrm{K}_I=\frac{\mathrm{I}_2}{\mathrm{I}_2}=\frac{Y_{21}Y_l}{Y_{11}(Y_{22}+Y_l)-Y_{12}Y_{21}}=-\frac{Z_{21}}{Z_{22}+Z_l}=\frac{H_{21}Y_l}{H_{22}+Y_l} \tag{5.18}$$

can be considered the *voltage* and *current gain* of the active two-port network, respectively.

In wide-band amplifiers, the amplifying devices (transistors, tubes, etc.), as a rule, ensure that the following inequalities hold (provided the load is chosen correctly):

$$Y_l \gg Y_{22};\quad Z_l \ll Z_{22} \tag{5.19}$$

Therefore, when roughly estimating the amplification ability of a two-port network, one can proceed from the following aproximate equalities:

$$|\mathrm{K}_E| \approx \left|\frac{Y_{21}}{Y_l}\right| \tag{5.20}$$

$$|\mathrm{K}_I| \approx \left|\frac{Z_{21}}{Z_{22}}\right| \tag{5.21}$$

From this it follows that the power gain (expressed in volt-amperes) is

$$\mathrm{K}_P=|\mathrm{K}_E||\mathrm{K}_I| \approx \to \approx \left|\frac{Y_{21}}{Y_l}\right|\left|\frac{Z_{21}}{Z_{22}}\right|=\frac{|Y_{21}|^2}{|Y_lY_{11}|} \tag{5.22}$$

(Here we use the relations between Z_{21}, Z_{22} and Y-parameters from Table 5.1.)

From (5.22) it is clear that the parameter Y_{21} (as well as Z_{21} or H_{21}) is of decisive importance in the amplification of power in active two-port networks. The physical meaning of this parameter is revealed later in the text.

When analyzing an active two-port network as an amplifier, special attention is given to such its parameters as the *input* and *output* impedances. Figure 5.6 shows a generalized circuit comprising a signal source E_s, an active two-port network, and a load impedance Z_l.

The input impedance (between the terminals *1* and *1′*) can easily be found by equations (5.4) in combination with (5.14).

Substituting $\mathbf{I}_2$ from (5.14) into the first equation of (5.4), we obtain

$$\mathbf{E}_1 = \mathbf{I}_1 \left(Z_{11} - \frac{Z_{12}Z_{21}}{Z'_{22}} \right) = \mathbf{I}_1 Z_{in}$$

whence

$$Z_{in} = Z_{11} - \frac{Z_{12}Z_{21}}{Z'_{22}} = Z_{11} - \frac{Z_{12}Z_{21}}{Z_{22}+Z_l} \tag{5.23}$$

By the output impedance of the two-port is meant the impedance between the terminals *2* and *2'* at $\mathbf{E}_s = 0$ (taking into account

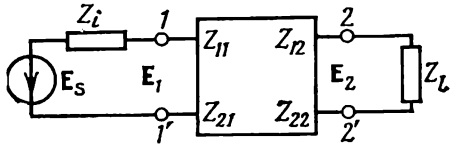


Fig. 5.6. Generalized circuit diagram of an active two-port network, taking into account the load and signal source parameters

the internal impedance Z_i of the signal source). In this case the impedance Z_i is regarded as a load.

By analogy with (5.23), on replacing Z_{11} by Z_{22} and Z_l by Z_i, we have

$$Z_{out} = Z_{22} - \frac{Z_{12}Z_{21}}{Z_{11}+Z_i} \tag{5.24}$$

If the internal impedance Z_i of the signal source is taken account of, the voltage gain will be $\mathbf{K}_E = \mathbf{E}_2/\mathbf{E}_s$. It can be found by adding Z_i to Z_{11} or H_{11} in (5.17):

$$\mathbf{K}_E = \frac{\mathbf{E}_2}{\mathbf{E}_s} = \frac{Z_{21}Z_l}{(Z_{11}+Z_i)(Z_{22}+Z_l)-Z_{12}Z_{21}} = -\frac{H_{21}}{(H_{11}+Z_i)(H_{22}+Y_l)-H_{12}H_{21}} \tag{5.25}$$

Using the Y-matrix, one can easily obtain the expression

$$\mathbf{K}_E = \frac{\mathbf{E}_2}{\mathbf{E}_s} = -\frac{Z_{in}}{Z_i+Z_{in}} \frac{Y_{21}}{Y_{22}+Y_l} \tag{5.26}$$

This expression coincides with the common definition of the *transfer function* of the linear two-port network.

From the above general relations it is clear that the structure of the transfer function of an active two-port network and the character of its dependence on frequency are determined by the frequency characteristics of the Z- or Y-parameters. In this respect there is no difference between active and passive linear two-port networks.

Setting $Z(\omega)$ or $Y(\omega)$ also unambiguously defines the time characteristics of a linear active circuit, i.e., its impulse response and transient response.

The dimensionless, and generally complex, function defined by formulas (5.17) and (5.18) is the most important characteristic of a two-port network. It is determined *under steady-state conditions with the two-port network being harmonically excited.*

The transfer function is conveniently represented in the form

$$\mathbf{K}(i\omega) = K(\omega)\, e^{i\varphi(\omega)} \tag{5.27}$$

The modulus $K(\omega)$ is sometimes called the *amplitude-frequency response* or simply *frequency response* of the two-port network. The argument $\varphi(\omega)$ of the transfer function is called the *phase-frequency response* or simply *phase response* of the two-port network.

For both active and passive linear circuits by the impulse response $g(t)$ of a circuit is meant the circuit response to the action of a unit impulse (delta function). The relation between $g(t)$ and $\mathbf{K}(i\omega)$ can easily be found by means of the Fourier integral.

If a unit impulse of electromotive force, having a spectral density equal to unity at all frequencies, is applied at the input of a two-port network, the spectral density of the output voltage is simply equal to $\mathbf{K}(i\omega)$. Therefore, the response to a unit impulse, i.e., the impulse response, of the circuit is easily determined by means of the inverse Fourier transform [see formula (2.49)] applied to the transfer function $\mathbf{K}(i\omega)$

$$v_{out}(t) = g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{K}(i\omega)\, e^{i\omega t}\, d\omega \tag{5.28}$$

The impulse response will later be denoted by the function $g(t)$ implying not only voltage but any other electrical quantity which is a response to the action of a delta function.

If the transfer ratio is written as $\mathbf{K}(p)$, i.e., as the Laplace transform of $g(t)$, expression (5.28) can be written* as the inverse Laplace transform

$$g(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \mathbf{K}(p)\, e^{pt}\, dp \tag{5.29}$$

The *transient response* $h(t)$ of the circuit is its response to a stimulus in the form of a "unit step". Since such a stimulus is the integral

* Here and elsewhere the transfer function of a given circuit, regarded as the Fourier or the Laplace transform of the impulse response $g(t)$, will differ only in the argument: $\mathbf{K}(i\omega)$ or $\mathbf{K}(p)$ (see Sec. 2.13).

of a unit impulse (i.e., of a delta function), there exists an integral relationship between $h(t)$ and $g(t)$:

$$h(t) = \int_0^t g(x)\,dx \tag{5.30}$$

In the following sections of this book, the impulse response $g(t)$ will most frequently be used for the analysis of signal transmission through radio circuits.

5.4. TRANSISTOR AMPLIFIER

Without going into the physics of the transistor shown schematically in Fig. 5.7, we shall treat it is a linear active two-port (or

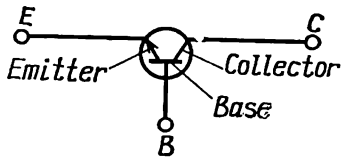


Fig. 5.7. Transistor as a three-terminal network

three-terminal) network which under conditions of low-level signal operation (and at comparatively low frequencies) feature the following properties.

(a) The emitter current is distributed between the base and the collector, the ratio of the collector current i_C to the emitter current i_E for a given transistor being practically constant and close to unity: $\alpha = i_C/i_E = 0.98$ to 0.998. Correspondingly, the base current is

$$i_B = (1 - \alpha)\, i_E \tag{5.31}$$

$$i_C/i_B = \alpha/(1 - \alpha) = \beta \gg 1 \tag{5.32}$$

(b) The emitter current is mainly determined by the base-to-emitter voltage and depends but slightly on the collector voltage. From this it follows that the collector current $i_C \approx \alpha i_E$ depends but slightly on the collector-to-emitter voltage.

These characteristic features of a transistor make it possible to construct its equivalent circuit as shown in Fig. 5.8*a*. In this diagram the dependent current source αi_E takes account of the effect of the emitter current on the collector circuit, while the resistances r_E, r_B and r_C are determined by the family of characteristics of the given transistor.

It should be noted that r_E, r_B and r_C are *differential resistances* for a-c current components whose amplitudes are sufficiently small

to justify the assumption of the linearity of the operating regions of the respective current-voltage characteristics of the transistor. In other words, the transistor operates under conditions of low-level signal amplification. The resistances r_E and r_B are relatively low (r_E — from several ohms to several tens of ohms, r_B — up to several

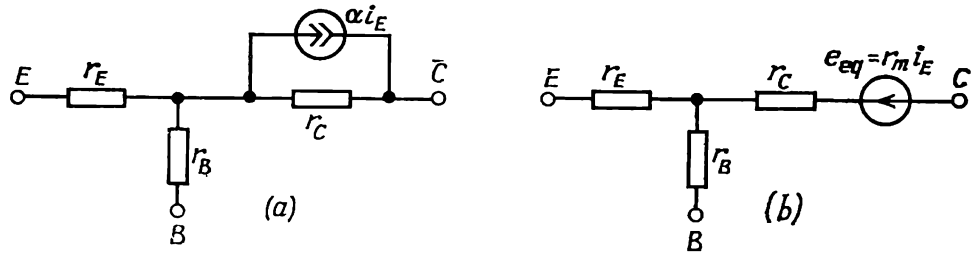


Fig. 5.8. Equivalent circuits of a transistor
(*a*) with a dependent current source; (*b*) with a dependent voltage source

hundred ohms), while the resistance r_C is very high (from hundreds of kilohms to several megohms).

In the equivalent circuit of Fig. 5.8*b*, the dependent current source αi_E with a shunt r_C is replaced by an equivalent voltage source with an internal resistance of r_C. The voltage of this source is

$$e_{eq} = r_C \alpha i_E = r_m i_E \tag{5.33}$$

where $r_m = \alpha r_C$.

The direction of e_{eq} is matched with the direction of the voltage which in the diagram of Fig. 5.8*a* produces the current αi_E when passing through r_C (the external collector circuit being open).

The amplification of signals in a transistor is due to the fact that the power dissipated in the high-value load resistor (in the collector circuit) by the a-c component of the collector current is much higher than the power of the signal source consumed in the base-to-emitter circuit to control the current. An increase in the power of the amplified signal occurs at the expense of the d-c source feeding the collector circuit.

In the text below, it is supposed that the amplifier input is fed with a harmonic oscillation, therefore, currents and voltages will be written in the form **I** and **E** implying the amplitudes (generally complex).

Depending on the choice of the input and output terminals, the following three amplifier circuit arrangements are possible: common-base, common-emitter and common-collector (Figs. 5.9, 5.10, and 5.11).

In the first arrangement (Fig. 5.9), the terminal B is common to both input and output circuits. In the circuit configurations shown

in Figs. 5.10 and 5.11, the common terminal is the terminal E and the terminal C, respectively.

Let us set up equations for voltages and currents in the above three circuit arrangements. To simplify the problem, let us disregard

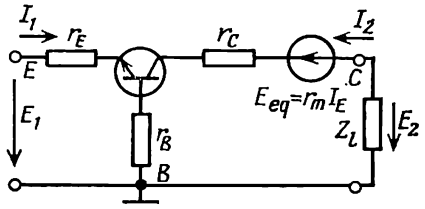


Fig. 5.9. Equivalent circuit of a common-base transistor amlpifier

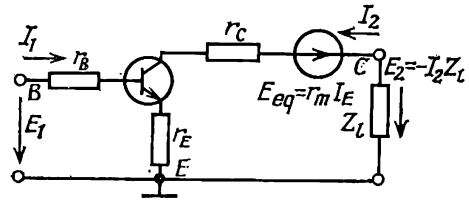


Fig. 5.10. Equivalent circuit of a common-emitter transistor amplifier

the interelectrode capacitances of the transistor, which is admissible at frequencies not exceeding a few megahertz. Since in this case we have purely ohmic resistances r_E, r_B, and r_C, the complex

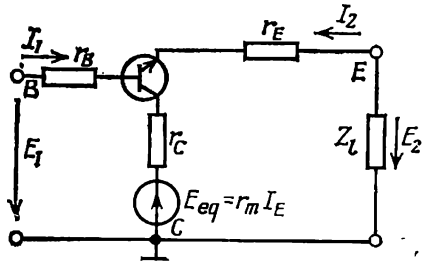


Fig. 5.11. Equivalent circuit of a common-collector transistor amplifier

amplitudes **I** and **E** can be replaced by their moduli I and E. Later on, when we shall take into consideration the impedances and admittances of the external circuits, we shall be able to transform to complex amplitudes.

For the common-base circuit arrangement (Fig. 5.9), the following two equations hold:

$$E_1 = (r_B + r_B)\, I_1 + r_B I_2, \quad E_2 - E_{eq} = r_B I_1 + (r_C + r_B)\, I_2$$

Substituting $E_{eq} = r_m I_E = r_m I_1$ into the second equation, we obtain the following system of equations:

$$E_1 = (r_E + r_B)\, I_1 + r_B I_2$$
$$E_2 = (r_m + r_B)\, I_1 + (r_C + r_B)\, I_2$$

Corresponding to these equations is a Z-matrix with the parameters

$$Z_{11}=r_E+r_B,\ Z_{12}=r_B$$
$$Z_{21}=r_m+r_B,\ Z_{22}=r_C+r_B$$

The common-emitter circuit configuration (Fig. 5.10) is described by the following equations:

$$E_1=(r_B+r_E)\,I_1+r_EI_2$$
$$E_2=(r_E-r_m)\,I_1+(r_C-r_m+r_E)\,I_2$$

Here we take into account that $E_{eq}=r_mI_E=r_m\,(I_1+I_2)$, the directions of E_{eq} and E_2 being coincident. In this case

$$Z_{11}=r_B+r_E,\ Z_{12}=r_E$$
$$Z_{21}=r_E-r_m,\ Z_{22}=r_C-r_m+r_E$$

Finally, the common-collector circuit arrangement (Fig. 5.11) has the following Z-parameters:

$$Z_{11}=r_B+r_C\approx r_C,\ Z_{12}=r_C-r_m$$
$$Z_{21}\approx r_C,\ Z_{22}=r_C-r_m+r_E$$

Thus, the elements of the Z-matrices for all the three transistor amplifier circuits can be expressed through the following physical parameters of the transistor:

$$r_E,\ r_B,\ r_C,\ \text{and}\ r_m=\alpha r_C$$

Using the obtained Z-parameters and Table 5.1, we can also find the Y- and H-parameters. Let us give these parameters for the most widely used common-emitter circuit arrangement. For this circuit configuration, the determinant of the Z-matrix is

$$\Delta Z=Z_{11}Z_{22}-Z_{12}Z_{21}=[(1-\alpha)\,r_B+r_E]\,r_C+r_Br_E$$
$$\approx[r_B+(1+\beta)\,r_E]\,r_C/(1+\beta) \qquad (5.34)$$

Here we have used the relation $\alpha=\beta/(1+\beta)$, stemming from expression (5.32), and omitted r_Br_E.

Thus, we have

$$Y_{11}=Z_{22}/\Delta Z=1/[r_B+(1+\beta)\,r_E]$$
$$Y_{12}=-Z_{12}/\Delta Z=-(1+\beta)\,r_E/[r_B+(1+\beta)\,r_E]\,r_C$$
$$Y_{21}=-Z_{21}/\Delta Z=\beta/[r_B+(1+\beta)\,r_E]$$
$$Y_{22}=Z_{11}/\Delta Z=[(1+\beta)\,(r_B+r_E)]/[r_B+(1+\beta)\,r_E]r_C$$
$$H_{11}=\Delta Z/Z_{22}=1/Y_{11}=r_B+(1+\beta)\,r_E$$
$$H_{12}=Z_{12}/Z_{22}\approx(1+\beta)\,r_E/r_C$$
$$H_{21}=-Z_{21}/Z_{22}\approx\beta$$
$$H_{22}=1/Z_{22}\approx(1+\beta)/r_C$$

These expressions are set up on condition that

$$r_E \ll (1 - \alpha) r_C$$

A characteristic operational feature of the transistor in the common-emitter circuit arrangement consists in controlling the collector current by acting on the base current. Furthermore, it is necessary to take into account the reverse action of the output voltage on the input circuit. These properties of the transistor are conveniently described by two-port equations (5.7). Therefore, the H-parameter matrix is now generally accepted in the theory and practice of transistor amplifiers.

In Sec. 5.3 it was shown that the amplifying ability of an active two-port network is mainly defined by the parameter H_{21} (respectively Y_{21} and Z_{21}). For a common-emitter amplifier circuit this parameter, as shown above, coincides with the coefficient β [see (5.32)]. This parameter enters into specifications of bipolar transistors and is designated by the symbol h_{21E}.

In accordance with the new designations, formulas (5.17) and (5.18) are written in the form

$$\mathbf{K}_E = \frac{\mathbf{E}_2}{\mathbf{E}_1} = -\frac{h_{21E}}{h_{11}(h_{22}+Y_l)-h_{21E}h_{12}} \tag{5.35}$$

$$\mathbf{K}_I = \frac{\mathbf{I}_2}{\mathbf{I}_1} = \frac{h_{21E}Y_l}{h_{22}+Y_l} \tag{5.36}$$

You will recall that h_{11} is the short-circuit input impedance of the base-emitter circuit, h_{12} is the open-circuit reverse-voltage transfer (feedback-voltage) ratio, and h_{22} is the open-circuit output admittance of the transistor.

Using the new designations, the second equation in (5.7) can be written in the following form:

$$\mathbf{I}_C = h_{21E}\mathbf{I}_B + h_{22}\mathbf{E}_2 = h_{21E}\mathbf{I}_B - h_{22}\mathbf{V}_{out} \tag{5.37}$$

where $\mathbf{V}_{out} = \mathbf{I}_C Z_l = -\mathbf{E}_2$ is the voltage across the load impedance $Z_l = 1/Y_l$.

Then, the base current $\mathbf{I}_B$ may be represented as the ratio $\mathbf{E}_1/Z_{in}$, where Z_{in} is the input impedance of the transistor (between the base and the emitter terminals) defined by formula (5.23). Thus, for purely ohmic resistances r_E, r_B, and r_C, when $Z_{in} = R_{in}$,

$$\begin{aligned}\mathbf{I}_C &\approx (h_{21E}/R_{in})\,\mathbf{E}_1 - h_{22}\mathbf{V}_{out} \\ &= S\mathbf{E}_1 - h_{22}\mathbf{V}_{out} = S\mathbf{E}_1 + h_{22}\mathbf{E}_2\end{aligned} \tag{5.37'}$$

The parameter $S = h_{21E}/R_{in}$ may be treated as the slope of the characteristic curve $i_C(v_{BE})$ at the point $v_{BE} = V_{BE0}$.

On the basis of expression (5.37') one can construct an equivalent circuit of the output circuit of an amplifier as shown in Fig. 5.12*a*.

The symbol R_i in Fig. 5.12*a* stands for the internal resistance of the current source. For the transistor in a common-emitter amplifier, $R_i = 1/h_{22}$.

From comparison of equation (5.37′) with (5.1) it follows that the parameter S coincides with the parameter Y_{21} (for the common-

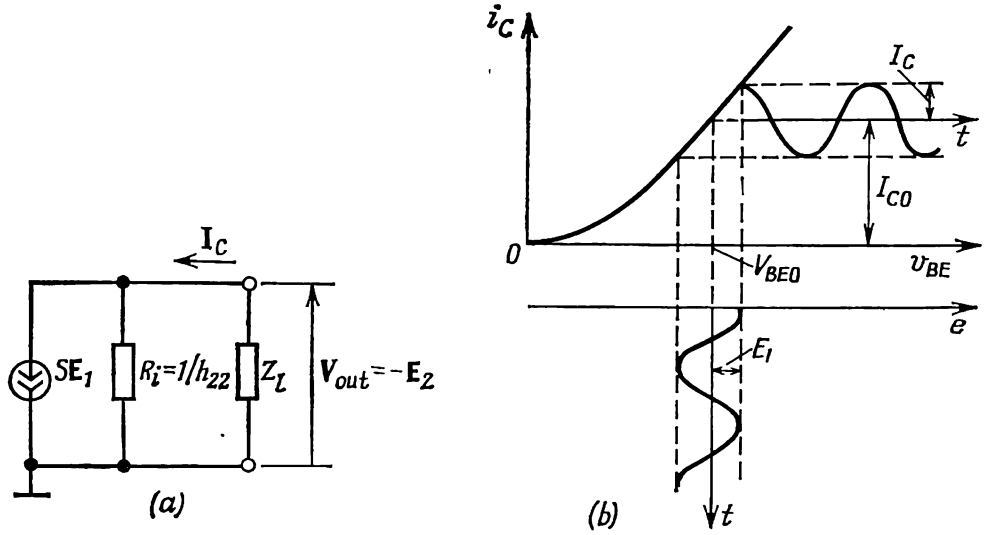


Fig. 5.12. (*a*) Equivalent circuit of the collector circuit and (*b*) linear operating conditions of a common-emitter transistor amplifier

emitter circuit arrangement). Substituting $I_C = -Y_l E_2$ into (5.37′) and dividing the obtained equation by $\mathbf{E}_1$, we get the following formula:

$$\mathbf{K}_E \approx -S/(h_{22} + Y_l) = -h_{21E}/R_{in}\,(h_{22} + Y_l) \qquad (5.35')$$

which differs from (5.35) only in appearance.

Where the admittance h_{22} is low as compared with the load admittance Y_l, one may use the approximate formulas

$$\mathbf{K}_E \approx -S/Y_l = -SZ_l \qquad (5.38)$$

$$\mathbf{K}_I \approx h_{21E} \qquad (5.39)$$

The low-level signal operation of a common-emitter transistor amplifier is illustrated in Fig. 5.12*b*. The amplitude of the a-c collector current $\mathbf{I}_C$ is much lower than the d-c current I_{C0} corresponding to the bias voltage V_{BE0}.

The three amplifiers illustrated in Figs. 5.9 through 5.11 essentially differ in properties.

Comparison of the common-emitter and common-base circuit arrangements leads to the following conclusions:

(a) both circuits are equivalent with respect to voltage amplification;

(b) in the common-emitter circuit, current is amplified by a factor of approximately h_{21E} [see. (5.39)] when $h_{22} \ll Y_l$, while in the common-base circuit, it is somewhat attenuated (insignificantly, because $I_2/I_1 = I_C/I_E = \alpha \approx 1$), hence, the power gain attainable with the common-emitter circuit is approximately h_{21E} times higher than that with the common-base circuit;

(c) an additional advantage of the common-emitter circuit consists in that it has a relatively high input impedance, this being explained by the fact that here the collector current is controlled by the base current which is much smaller than the emitter current;

(d) in the common-base circuit, the output voltage is in phase with the input voltage, while in the common-emitter circuit these voltages are in antiphase.

The common-collector amplifier circuit occupies a special position. The voltage across the terminals B and E (Fig. 5.11) is the difference between the voltages E_1 and E_2. The voltage drop across the load impedance Z_l due to the current I_E is always lower than E_1, therefore, the voltage gain in the common-collector circuit is always less than unity, while the current gain is close to h_{21E}. Consequently, the common-collector amplifier may be regarded as a *constant-voltage current amplifier*. The load impedance Z_l inserted into the emitter circuit may be very low, much lower than in cases where it is inserted into the collector circuit (as in the circuits of Figs. 5.9 and 5.10). This is of great advantage, since in this case the influence of the load capacitance, that shunts the amplifier output, on the frequency response of the amplifier is minimized. Of no less importance is also the fact that the output voltage relative to the common point (ground) is in phase with the input voltage (i.e., they are of the same polarity). Thus, the common-collector amplifiers "follows" the signal changing neither its shape, amplitude (voltage), or polarity, but providing a match between the high-impedance input and the low-impedance load Z_l. Therefore, the common-collector amplifier circuit is often called *emitter follower*. Owing to these properties, the emitter follower has found wide application as dependent *voltage-controlled voltage source* (in the ideal case, such a source must have an infinitely high input impedance and a zero output impedance).

From this point of view, the common-emitter amplifier circuit having a comparatively high input impedance and a very high output impedance may be regarded as dependent *voltage-controlled current source* (in the ideal case, both impedances must be infinitely high).

Finally, the common-base amplifier circuit having a low input

impedance and a high output impedance approaches a dependent *current-controlled current source.*

In conclusion, it should be noted that the above-described equivalent circuits of transistor amplifiers are valid at frequencies not exceeding several megahertz. At higher frequencies, it is necessary to take into account the dependence of the factor α on frequency and also, the effect of some interelectrode capacitances which have not been taken into account when constructing the equivalent circuits (Figs. 5.8 through 5.12). These questions are treated in textbooks on amplifier systems.

5.5. ELECTRON-TUBE AMPLIFIER

Figure 5.13*a* shows a simple amplifier circuit using a pentode. With a low-level signal (linear amplification conditions), the anode

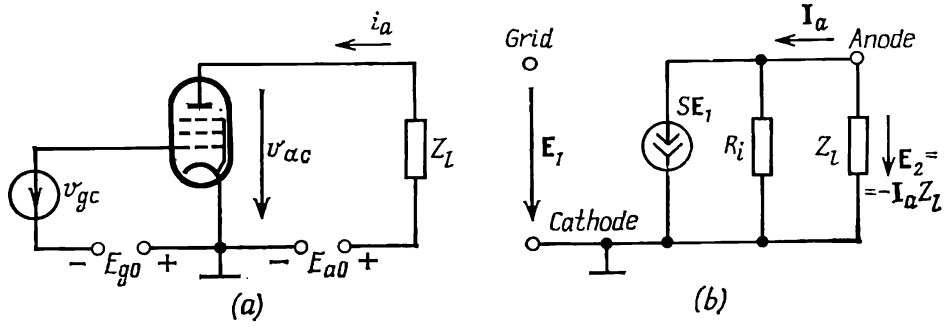


Fig. 5.13. (*a*) Simple pentode amplifier and (*b*) equivalent circuit of the anode circuit

current and the grid-to-cathode and anode-to-cathode voltages are related by

$$i_a = Sv_{gc} + (1/R_i)\, v_{ac} = S\,(v_{gc} + Dv_{ac}) \tag{5.40}$$

where

$$S = \frac{di_a}{dv_{gc}} \text{ when } v_{gc} = E_{g0},\ v_{ac} = E_{a0}$$

$$\frac{1}{R_i} = \frac{di_a}{dv_{ac}} \text{ when } v_{gc} = E_{g0},\ v_{ac} = E_{a0}$$

$D = 1/SR_i$ is the penetration factor (this relation holds for operation without grid current).

The transconductance S, i.e., the slope of the characteristic curve $i_a\,(v_{gc})$, and the internal resistance R_i of the pentode are differential parameters defined with insignificant deviations of the

current i_a from its initial value i_{a0} at the operating point on the current-voltage curve of the pentode.

The plus sign before the second term in expression (5.40) is due to the fact that in this case v_{ac} is regarded as the voltage of an independent source.

For the grid current we may set up an expression similar to (5.40):

$$i_g = (1/R_{gc})\, v_{gc} + S_{ga} v_{ac} \qquad (5.40')$$

where R_{gc} is the grid-to-cathode resistance and S_{ga} is the slope of the characteristic curve $i_g\,(v_{ac})$.

Proceeding now to complex amplitudes and considering the common equivalent circuit of an active two-port network (Fig. 5.2), let us replace v_{gc} by the amplitude $\mathbf{E}_1$ of the input harmonic signal, the grid current i_g, by the amplitude $\mathbf{I}_1$, and the current i_a, by the amplitude $\mathbf{I}_2 = \mathbf{I}_a$. As in the preceding section (see Fig. 5.12*a*), we assume that $\mathbf{E}_2 = -\mathbf{I}_a Z_l = -\mathbf{V}_{out}$. Then equations (5.1) and (5.3) will be written in the form

$$\mathbf{I}_1 = \frac{1}{R_{gc}}\mathbf{E}_1 + S_{ga}\mathbf{E}_2, \quad \mathbf{I}_2 = S\mathbf{E}_1 + \frac{1}{R_i}\mathbf{E}_2$$

$$[Y] = \begin{bmatrix} 1/R_{gc} & S_{ga} \\ S & 1/R_i \end{bmatrix} \qquad (5.41)$$

When small signals are amplified, the operating point on the characteristic curve $i_a\,(v_{gc})$ is, as a rule, within the region of the negative v_{gc}. In this case, there is no grid current, the input grid-to-anode conductance is practically zero ($R_{gc} \to \infty$), and the admittance matrix takes the form

$$[Y] = \begin{bmatrix} 0 & 0 \\ S & 1/R_i \end{bmatrix} \qquad (5.41')$$

Thus, $Y_{11} = Y_{12} = 0$; $Y_{21} = S$; $Y_{22} = 1/R_i$.

The matrix (5.41′) corresponds to the equivalent circuit of the two-port (three-terminal) network shown in Fig. 5.13*b*.

Using formula (5.17), let us find the voltage gain:

$$\mathbf{K}_E = \frac{\mathbf{E}_2}{\mathbf{E}_1} = -\frac{S}{1/R_i + 1/Z_l} = -\frac{SR_iZ_l}{R_i + Z_l} = -\frac{\mu Z_l}{R_i + Z_l} \qquad (5.42)$$

where $\mu = SR_i = 1/D$ is the amplification factor of the tube.

From expression (5.42) it is clear that the greater the ratio Z_l/R_i, the fuller the use of the amplifying ability μ of the tube. Under no-load conditions ($Z_l \to \infty$) the amplification factor of the amplifier stage is

$$\max K_E \to -\mu = -SR_i = -S/G_i$$

Along with the above-considered common-cathode circuit, common-grid and common-anode circuits are also possible. These three circuits are analogs of the respective transistor amplifier circuits (common-emitter, common-base, and common-collector).

5.6. APERIODIC AMPLIFIER

Figure 5.14*a* shows an equivalent circuit of a simple aperiodic resistance-coupled amplifier. The amplifying element is shown in the form of a dependent current force $S\mathrm{E}_1$ with internal conductance $G_i = 1/R_i$. The capacitance C_0 includes the interelectrode capa-

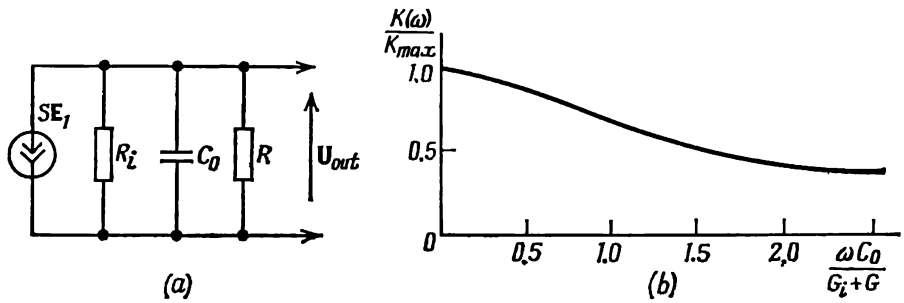


Fig. 5.14. Aperiodic common-emitter transistor amplifier
(*a*) equivalent circuit of the collector circuit; (*b*) amplitude-frequency characteristic

citance of the active element and also the capacitance of the external network shunting the load resistor $R = 1/G$. The circuit of Fig. 5.14*a* is a generalization applicable to any active element.

In the case of a transistor amplifier, the transconductance S implies the quantity h_{21E}/R_{in} (see Section 5.4) and G_i, the parameter h_{22}.

Substituting G_i for h_{22} and also $Y_l = 1/R + i\omega C_0 = G + i\omega C_0$ in formula (5.35′), we obtain the transfer function of the amplifier

$$\mathbf{K}_E(i\omega) = -\frac{S}{G_i+G+i\omega C_0} = -\frac{S/(G_i+G)}{1+i\omega C_0/(G_i+G)} = -\frac{K_{\max}}{1+i\omega C_0/(G_i+G)} \tag{5.43}$$

where

$$K_{\max} = S/(G_i + G) \tag{5.44}$$

is the maximum amplification factor (with $\omega \to 0$).

The modulus of the transfer function

$$K(\omega) = K_{\max}/\sqrt{1+[\omega C_0/(G_i+G)]^2} \tag{5.45}$$

Letting the frequency ω change, we obtain the amplitude-frequency characteristic shown in Fig. 5.14*b*.

In aperiodic amplifiers built around pentodes, use is often made of the circuit shown in Fig. 5.15*a*. This circuit differs from the one shown in Fig. 5.13*a* only in that it contains an additional network R_1, C_1, which is intended to protect the output circuits of the amplifier from the d-c supply voltage E_{a0}. As a rule, the resistor R_1 has a high value, much higher than that of R, which is selected so as to

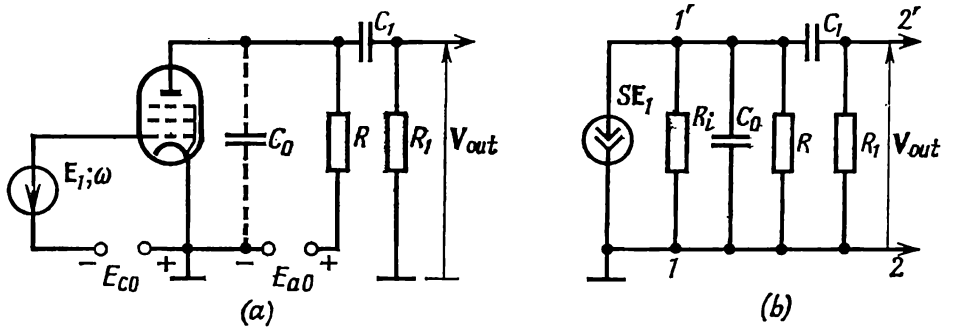


Fig. 5.15. (*a*) Aperiodic vacuum-tube amplifier with a separation circuit, R_1C_1, at the output and (*b*) equivalent circuit of the anode circuit

ensure that the amplifier has a sufficiently wide passband. An equivalent circuit of the anode circuit of the amplifier is shown in Fig. 5.15*b*.

First, let us find the voltage across the terminals *1* and *1'* (Fig. 5.15*b*) due to the current $S\mathrm{E}_1$. It is clear that

$$\mathrm{E}_{1\text{-}1'} = \frac{S\mathrm{E}_1}{G_i + i\omega C_0 + G + \dfrac{1}{R_1 + 1/i\omega C_1}} = \frac{S\mathrm{E}_1}{G_i + i\omega C_0 + G + \dfrac{i\omega C_1 G_1}{G_1 + i\omega C_1}} \tag{5.46}$$

In this case, the output voltage of the amplifier, i.e., the voltage across the terminals *2* and *2'*, is

$$\mathrm{V}_{out} = \mathrm{E}_{1\text{-}1'} \frac{R_1}{R_1 + 1/i\omega C_1} = \mathrm{E}_{1\text{-}1'} \mathrm{K}_1$$

where

$$\mathrm{K}_1 = \frac{R_1}{R_1 + 1/i\omega C_1} = i \frac{\omega C_1}{G_1 + i\omega C_1} \tag{5.47}$$

is the transfer function of the separation circuit R_1, C_1.

On the basis of expressions (5.46) and (5.47) we obtain

$$\mathrm{K}_E(i\omega) = -\frac{\mathrm{V}_{out}}{\mathrm{E}_1} = \frac{\mathrm{E}_{1\text{-}1'}\mathrm{K}_1}{\mathrm{E}_1} = -\frac{S}{G_i + i\omega C_0 + G + i\,[G_1\omega C_1/(G_1 + i\omega C_1)]} \times \frac{i\omega C_1}{[(G_1 + i\omega C_1)} = -\frac{S}{G_\Sigma - i\,[(G_i + G)G_1/\omega C_1 - \omega C_0]} \tag{5.48}$$

where $G_\Sigma = G_i + G + G_1$ (the term G_1C_0/C_1^2 is omitted, because $C_0/C_1 \ll 1$).

Let us discuss the frequency characteristic of the amplifier. In the region adjoining $\omega = 0$, i.e., at frequencies where the reactance $1/\omega C_1$ of the blocking capacitor is higher than, or commensurable with, the resistance R_1, the influence of the susceptance $i\omega C_0$, as well as of the susceptance $1/(R_1 + 1/i\omega C_1)$ in (5.46), may be disregarded so that expression (5.48) becomes

$$\mathbf{K}_E(i\omega) \approx -\frac{S}{G_i+G}\,\frac{i\omega C_1}{G_1+i\omega C_1}$$

The modulus of this expression is

$$K_E(\omega) = \frac{S}{G_i+G}\,\frac{\omega R_1 C_1}{\sqrt{1+(\omega R_1 C_1)^2}},\ \omega \leqslant \frac{1}{R_1 C_1} \qquad (5.49)$$

In the central region of this characteristic, at frequencies where $1/\omega C_1 \ll R_1$, while the susceptance ωC_0 is still negligible compared with the conductance G, formula (5.49) is further simplified:

$$K_E(\omega) = \frac{S}{G_\Sigma} = K_{\max} \text{ when } \frac{1}{R_1 C_1} \ll \omega \ll \frac{G_\Sigma}{C_0} \qquad (5.49')$$

Finally, at comparatively high frequencies, where the susceptance ωC_0 is commensurable with G_Σ, we obtain

$$K_E(\omega) = \frac{S}{\sqrt{G_\Sigma^2+(\omega C_0)^2}} = \frac{K_{\max}}{\sqrt{1+(\omega C_0/G_\Sigma)^2}},\quad \omega C_0 \gg G_\Sigma \qquad (5.49'')$$

Letting the frequency ω change, we obtain the frequency characteristic shown in Fig. 5.16*a*.

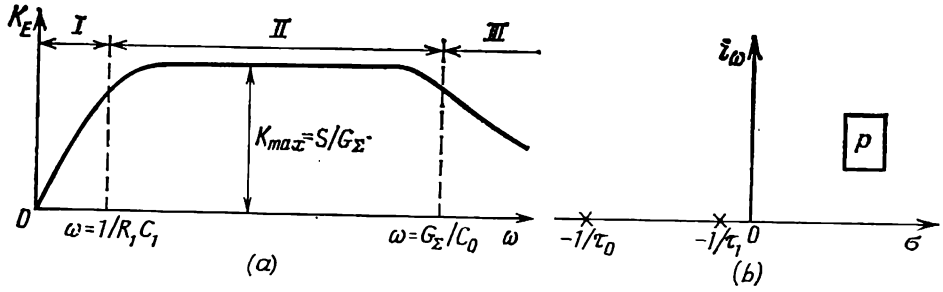


Fig. 5.16. (*a*) Frequency characteristic of the amplifier shown in Fig. 5.15 and (*b*) the poles of its transfer function

Usually, the following inequalities are satisfied: $R_i \gg R$; $R_1 \gg R$. Therefore, we may ignore the conductances G_i and G_1 and consider that $G_\Sigma \approx G \approx 1/R$.

In this case, expression (5.48) is simplified:

$$\mathbf{K}_E(i\omega) \approx -\frac{SR}{1-i\,(1/\omega R_1 C_1 - \omega R C_0)} \approx -\frac{K_{\max}}{1+i\omega\tau_0+1/i\omega\tau_1} \qquad (5.50)$$

where $\tau_0 = RC_0$ is the time constant of the circuit R, C_0 and $\tau_1 = R_1C_1$, the time constant of the separation circuit R_1, C_1;

$$K_{max} \approx -SR \tag{5.51}$$

is the amplification factor (approximate) in the frequency range $1/R_1C_1 < \omega < 1/RC_0$.

Transforming to the complex variable $p = \sigma + i\omega$, we write expression (5.50) in the form

$$\mathbf{K}_E(p) = -\frac{K_{max}}{1 + p\tau_0 + 1/p\tau_1} = -\frac{K_{max}p}{\tau_0 p^2 + p + 1/\tau_1} \tag{5.52}$$

The poles of the transfer function $\mathbf{K}(p)$ (i.e., the roots of the equation $\tau_0 p^2 + p + 1/\tau_1 = 0$) are

$$p_{1,2} = -\frac{1}{2\tau_0} \pm \sqrt{\frac{1}{4\tau_0^2} - \frac{1}{\tau_0\tau_1}} = -\frac{1}{2\tau_0}\left(1 \mp \sqrt{1 - \frac{4\tau_0}{\tau_1}}\right)$$

Since $\tau_0 \ll \tau_1$, we have

$$\left.\begin{aligned} p_1 &\approx -\frac{1}{2\tau_0}\left[1 - \left(1 - \frac{2\tau_0}{\tau_1}\right)\right] \approx -\frac{1}{\tau_1} \\ p_2 &\approx -\frac{1}{2\tau_0}\left[1 + \left(1 - \frac{2\tau_0}{\tau_1}\right)\right] \approx -\frac{1}{\tau_0} \end{aligned}\right\} \tag{5.53}$$

The disposition of the poles p_1 and p_2 on the plane p is shown in Fig. 5.16*b*.

It should be noted that as $G_1 \to \infty$ (with a finite and constant R_1) formula (5.48) reduces to

$$\mathbf{K}_E(i\omega) = -\frac{S}{G_\Sigma + i\omega C_0} = -\frac{S}{G_\Sigma}\left(\frac{1}{1 + i\omega C_0/G_\Sigma}\right) = -\frac{K_{max}}{1 + i\omega C_0/G_\Sigma} \tag{5.51'}$$

which coincides with (5.43) when $G_i + G$ in the latter is replaced by G_Σ. In this case, the only pole of the transfer function

$$\mathbf{K}_E(p) = \frac{K_{max}}{1 + C_0 p/G_\Sigma} \tag{5.52'}$$

will be at the point $p_1 = -G_\Sigma/C_0$.

Relations (5.44) through (5.53) will be used in the next chapter when analyzing the transmission of signals through an aperiodic amplifier. In the same chapter, we shall consider the impulse response of the given amplifier.

5.7. RESONANCE AMPLIFIER

A single-tuned common-emitter transistor amplifier shown in Fig. 5.17*a* differs from those discussed in the preceding section only in the load circuit. In the given case, the load is the resistor R_{sh} that shunts a parallel resonant circuit. As a rule, the power

losses in the inductance coil L and capacitor C may be disregarded, because they are low in comparison with the power dissipated in the resistor R_{sh}. In this case, the load admittance (across the terminals *1* and *2*) is given by

$$Y_l = G_{sh} + i\omega C + \frac{1}{i\omega L} = G_{sh}\left[1 + iR_{sh}\left(\omega C - \frac{1}{\omega L}\right)\right] \quad (5.54)$$

Using the relations $\omega_r = 1/\sqrt{LC}$ and $\rho = \sqrt{L/C} = \omega_r L = 1/\omega_r C$, where ω_r is the resonance frequency and ρ is the characteristic

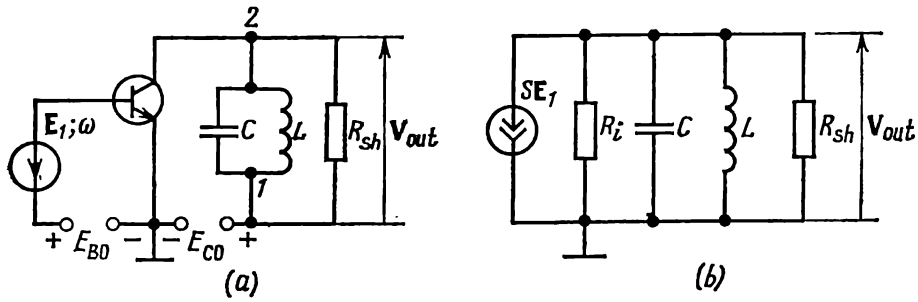


Fig. 5.17. (*a*) Resonance amplifier and (*b*) equivalent circuit of the collector circuit

impedance of the resonant circuit, expression (5.54) may be reduced to the form

$$Y_l = G_{sh}\left[1 + i\frac{R_{sh}}{\rho}\left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega}\right)\right] = G_{sh}\left[1 + iQ\left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega}\right)\right] \quad (5.55)$$

At a sufficiently high quality factor of the resonant circuit, $Q = R_{sh}/\rho$, most important is the value of the transfer function of the amplifier near the resonance frequency, i.e., where the deviations of ω from ω_r are small. Therefore, we may impose the condition of small relative detuning:

$$\frac{|\omega - \omega_r|}{\omega_r} = \frac{|\Delta\omega|}{\omega_r} \ll 1 \quad (5.56)$$

Then, substituting $\omega = \omega_r + \Delta\omega$, we obtain

$$\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega} = \left(1 + \frac{\Delta\omega}{\omega_r}\right) - \frac{1}{1 + \Delta\omega/\omega_r} \approx \frac{2\Delta\omega}{\omega_r} \quad (5.57)$$

and expression (5.55) may be written in the form

$$Y_l = G_{sh}\left(1 + i\frac{2\Delta\omega}{\omega_r}Q\right) = G_{sh}(1 + ia) \quad (5.58)$$

where

$$a = (2\Delta\omega/\omega_r)\,Q \quad (5.59)$$

has the meaning of *generalized detuning* of the resonant circuit.

Now let us set up an expression for the transfer function (with respect to voltage) of the resonance amplifier. Using the equivalent circuit shown in Fig. 5.17*b* and by analogy with formula (5.43), we may write

$$\mathbf{K}_E(i\omega) = -\frac{S}{G_i + Y_l} = -\frac{S}{G_i + G_{sh} + i\omega C + 1/i\omega L} \tag{5.60}$$

Substituting into this expression Y_l defined by formula (5.58), we may express the transfer function through the generalized detuning a as follows:

$$\mathbf{K}_E(i\omega) = \mathbf{K}(ia) = -S/(G_i + G_{sh} + iaG_{sh}) \tag{5.60'}$$

It should be noted that at resonance ($a = 0$) the amplification factor

$$K_{\max} = S/(G_i + G_{sh}) \tag{5.61}$$

Therefore, (5.60′) may be written in a somewhat different form

$$\mathbf{K}(ia) = -\frac{K_{\max}}{1 + iaG_{sh}/(G_i + G_{sh})} \tag{5.62}$$

Introducing a new symbol for the generalized detuning (taking into account the influence of the internal conductance G_i)

$$a_{eq} = \frac{aG_{sh}}{G_i + G_{sh}} = \frac{2\Delta\omega}{\omega_r} Q_{eq} \tag{5.63}$$

we come to the final expression for the transfer function of the amplifier

$$\mathbf{K}_1(ia_{eq}) = -\frac{K_{\max}}{1 + ia_{eq}} = \frac{K_{\max}}{\sqrt{1 + a_{eq}^2}} e^{i[\varphi(a_{eq}) + \pi]} \tag{5.64}$$

The term π in the exponent takes account of the minus sign on the right-hand side of (5.62).

The moduli ratio

$$n(a_{eq}) = \frac{K_1(a_{eq})}{K_{\max}} = \frac{1}{\sqrt{1 + a_{eq}^2}} \tag{5.65}$$

may be regarded as the normalized amplitude-frequency response of the single-tuned amplifier, while the argument

$$\varphi(a_{eq}) = -\arctan a_{eq} \tag{5.66}$$

may be regarded as the phase-frequency response (without taking into consideration the shift π which is independent of frequency).

The characteristics $n(a_{eq})$ and $\varphi(a_{eq})$ (Fig. 5.18) do not differ in any way from the characteristics of a passive oscillatory circuit having the same Q-factor.

The relative passband of a resonance amplifier, defined by the attenuation of the amplitude at the boundary of the band to $1/\sqrt{2}$ of the maximum level (at $a_{eq}=0$) and expressed through the generalized detuning a_{eq}, is equal to 2 (see Fig. 5.18). To transform from the dimensionless relative passband 2 to the dimensional passband $2\Delta\omega_0$, let us set in (5.63) $|a_{eq}|=1$ and $|\Delta\omega|=\Delta\omega_0$. In this case, the passband will be

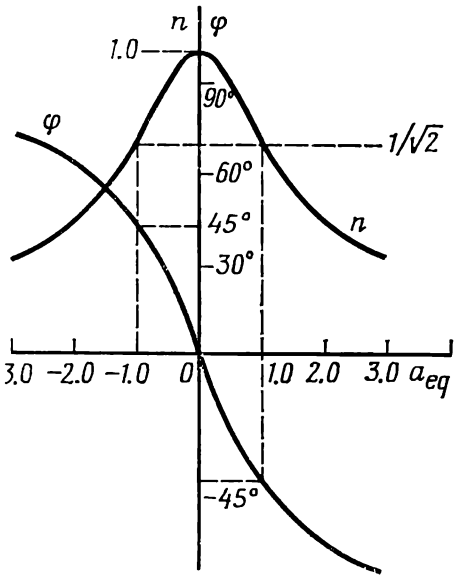


Fig. 5.18. Amplitude-frequency and phase-frequency characteristics of a single-tuned resonance amplifier

$$2\Delta\omega_0=\omega_0/Q_{eq} \qquad (5.67)$$

where Q_{eq} [as it follows from (5.63)] is the Q-factor of the loaded resonant circuit.

In many practical cases the internal conductance G_i of an amplifier is low as compared with the load conductance G_{sh} (respectively, $R_i \gg R_{sh}$). For rough calculations, formulas (5.61) and (5.62) may be simplified as follows:

$$K_{\max}\approx\frac{S}{G_{sh}}, \quad \mathbf{K}(ia)\approx-\frac{S}{G_{sh}}\frac{1}{1+ia} \qquad (5.68)$$

Where the amplifier load is expressed through the resistance r_{int} introduced into the circuit, the resistance R_{sh} in the preceding formulas must be replaced by the equivalent impedance of the parallel resonant circuit at resonance:

$$Z_{eq.\,r}=\rho^2/r_{int}=L/Cr_{int}$$

In this case

$$K_{\max}\approx SZ_{eq.\,r}, \quad \mathbf{K}(ia)\approx-SZ_{eq.\,r}\left(\frac{1}{1+ia}\right) \qquad (5.69)$$

5.8. FEEDBACK IN AN ACTIVE TWO-PORT NETWORK

In Secs. 5.4 through 5.7, when analyzing linear amplifiers on the basis of the parameter matrices of equivalent two-port networks, attention was mainly given to the parameters Y_{21}, Z_{21}, and H_{21}, since it is exactly these parameters that define the amplifying

ability of an active two-port network. In real two-port networks, which are not completely unidirectional, one has to take into account the effect of the output oscillation on the amplifier input.

Let the voltage and current at the output of the amplifier, whose equivalent circuit is shown in Fig. 5.19, be $\mathbf{E}_2$ and $\mathbf{I}_2$ (under operating conditions).

Regarding these quantities as being the result of an external stimulus at the amplifier output, we can determine $\mathbf{I}_1'$ and $\mathbf{E}_1'$ at the input by means of the equivalent circuit (Fig. 5.19).

In this circuit, the terminals *1* and *1'* to which the input signal source is connected are conditionally short-circuited, and by the

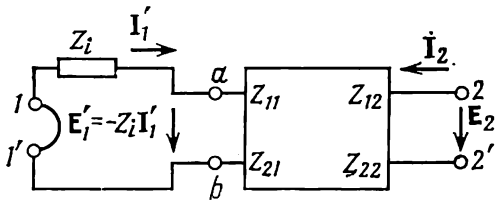


Fig. 5.19. Determining the reaction of an amplifier

voltage acting across the terminals *a* and *b* is ment $\mathbf{E}_1' = -Z_i\mathbf{I}_1'$, i.e., the voltage drop across the internal impedance Z_i of the current source due to the current $\mathbf{I}_1'$.

Equations (5.4) in terms of Fig. 5.19 are written in the form $\mathbf{E}_1' = -Z_i\mathbf{I}_1' = Z_{11}\mathbf{I}_1' + Z_{12}\mathbf{I}_2$, $\mathbf{E}_2 = Z_{21}\mathbf{I}_1' + Z_{22}\mathbf{I}_2$, whence we may easily obtain the following relation:

$$\frac{\mathbf{E}_1'}{\mathbf{E}_2} = Z_{12}\frac{Z_i}{Z_{11}Z_{22} - Z_{12}Z_{21} + Z_{22}Z_i} = Z_{12}\frac{Z_i}{\Delta Z + Z_{22}Z_i} \qquad (5.70)$$

The voltage $\mathbf{E}_1'$ is often called *feedback voltage*, Z_{12} being the feedback element. When representing an equivalent circuit of a two-port network by means of a Y- or an H-matrix, the parameters Y_{12} and H_{12} are the feedback elements, respectively.

The above-discussed feedback, which is conditioned by the physical parameters of the amplifying element, may be called *internal feedback*. As a rule, it results in undesirable phenomena—dependence of the input circuit parameters of the amplifier on the load elements, danger of instability under some conditions, etc.

Let us discuss the basic concepts concerning the application of *external feedback* in amplifiers. The simplest way to effect such a feedback is to connect the amplifier output to the input through a one-port network (Fig. 5.20).

With the output being connected to the input through a feedback one-port network Y_{fb} as shown in Fig. 5.20*a*, the main two-port network is preferably described by means of a Y-matrix. Taking

into account the obvious equality $\mathbf{I}_1 = \mathbf{I}_1' - Y_{fb}(\mathbf{E}_2 - \mathbf{E}_1)$, as well as the relations between $\mathbf{I}_1'$, $\mathbf{I}_2'$ and $\mathbf{E}_1$, $\mathbf{E}_2$ in the form of equations (5.1), we arrive at a new system of equations

$$\left.\begin{aligned}\mathbf{I}_1 &= (Y_{11}+Y_{fb})\,\mathbf{E}_1 + (Y_{12}-Y_{fb})\,\mathbf{E}_2\\ \mathbf{I}_2 &= (Y_{21}-Y_{fb})\,\mathbf{E}_1 + (Y_{22}+Y_{fb})\,\mathbf{E}_2\end{aligned}\right\} \tag{5.71}$$

Thus, the two-port network with feedback, shown in Fig. 5.20*a*, is represented by the admittance matrix

$$[Y]' = \begin{bmatrix} Y_{11}+Y_{fb} & Y_{12}-Y_{fb} \\ Y_{21}-Y_{fb} & Y_{22}+Y_{fb} \end{bmatrix} \tag{5.72}$$

from which it follows that the introduction of the one-port network Y_{fb} changes all the elements of the matrix, including the feedback element ($Y_{12} - Y_{fb}$ instead of Y_{12}).

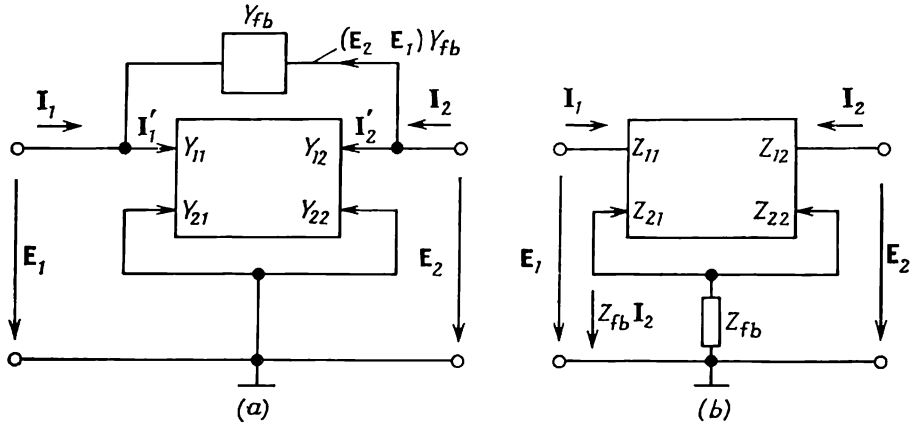


Fig. 5.20. Amplifier circuits with (*a*) voltage feedback and (*b*) current feedback

In a similar way, we may show that the connection of the one-port network Y_{fb} as shown in Fig. 5.20*b* yields the matrix

$$[Z]' = \begin{bmatrix} Z_{11}+Z_{fb} & Z_{12}+Z_{fb} \\ Z_{21}+Z_{fb} & Z_{22}+Z_{fb} \end{bmatrix} \tag{5.73}$$

In the circuit of Fig. 5.20*a*, the additional current fed from the output to the input via the feedback network is equal to $(\mathbf{E}_2 - \mathbf{E}_1)\,Y_{fb}$. Since in amplifiers we usually have $\mathbf{E}_2 \gg \mathbf{E}_1$, the value of this current is approximately equal to $\mathbf{E}_2 Y_{fb}$, i.e., it is proportional to the output voltage. Therefore, the amplifier shown in Fig. 5.20*a* may be called an amplifier with *voltage feedback*. In the circuit of Fig. 5.20*b*, where the feedback voltage is proportional to the output current, we have *current feedback*. It is obvious that one can effect combined voltage-and-current feedback.

Two types of feedback are distinguished: *negative* and *positive*. If the introduction of feedback increases the gain (absolute) of the circuit, feedback is positive, otherwise it is negative.

Let us explain the application of expressions (5.71) and (5.72) to a comon-emitter transistor amplifier circuit with $Y_{fb}=1/R_{fb}$ (Fig. 5.21).

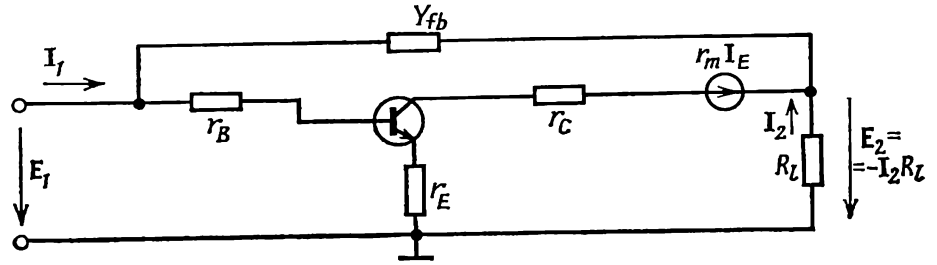


Fig. 5.21. Example of an equivalent circuit of a common-emitter amplifier with external feedback

Using formula (5.17) with Y_{21} replaced by $Y_{21}-Y_{fb}$ and Y_{22}, by $Y_{22}+Y_{fb}$ [see (5.72)], we find the voltage gain to be

$$\mathbf{K}_E=-\frac{Y_{21}-Y_{fb}}{(Y_{22}+Y_{fb})+Y_l} \tag{5.74}$$

In this expression, the admittances Y_{21} and Y_{22}, expressed through transistor parameters h_{21E}, r_B, r_E, and r_C (see Sec. 5.4) as

$$Y_{21}\approx\frac{h_{21E}}{r_B+(1+h_{21E})\,r_E},\quad Y_{22}=\frac{(1+h_{21E})\,(r_B+r_E)}{[r_B+(1+h_{21E})\,r_E]\,r_C} \tag{5.75}$$

are both real and positive numbers. The same is true for $Y_{fb}=1/R_{fb}$, $Y_l=1/R_l$. It is clear that subtracting Y_{fb} from the numerator of the fraction in (5.74) and adding it to the denominator reduce the magnitude of the amplification factor, i.e., in the given case feedback is *negative*. This is due to the fact that the input and output voltages in the resistance-coupled common-emitter amplifier are in antiphase (see Sec. 5.4). The current through R_{fb} ($\approx \mathbf{E}_2/R_{fb}=-\mathbf{I}_2R_l/R_{fb}$) flowing from the output to the input reduces the current I_B and hence, $\mathbf{E}_2$.

It can be shown that a similar connection of the one-port network $Y_{fb}=1/R_{fb}$ into the common-base amplifier circuit where the voltages $\mathbf{E}_2$ and $\mathbf{E}_1$ are in phase, yields *positive* feedback.

Figure 5.22 shows a structural diagram of an amplifier with external voltage feedback effected by means of an auxiliary two-port network $\mathbf{K}_{fb}(i\omega)$.

Both the amplifier $\mathbf{K}_a(i\omega)$ and the two-port network $\mathbf{K}_{fb}(i\omega)$ are supposed to be completely unidirectional. Such supposition makes sense where the input impedance of the two-port network $\mathbf{K}_{fb}(i\omega)$ is sufficiently high not to load the amplifier $\mathbf{K}_a(i\omega)$ and the output impedance of the two-port network $\mathbf{K}_{fb}(i\omega)$ is suffici-

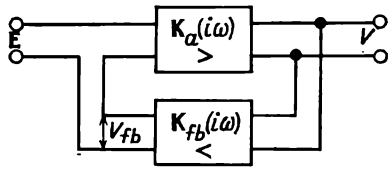


Fig. 5.22. Block diagram of an amplifier with feedback

ently low compared to the input impedance of the amplifier $\mathbf{K}_a(i\omega)$. Under these assumptions, the transfer function of the system as a whole

$$\mathbf{K}_0(i\omega) = \mathbf{V}/\mathbf{E} \tag{5.76}$$

can be found by means of the following obvious relations. The voltage at the output of the two-port feedback network is

$$\mathbf{V}_{fb} = \mathbf{K}_{fb}(i\omega)\,\mathbf{V} \tag{5.77}$$

The input voltage of the amplifier $\mathbf{K}_a(i\omega)$ is equal to the sum of the input e.m.f. $\mathbf{E}$ and the feedback voltage $\mathbf{V}_{fb}$.

Therefore, the output voltage of the entire circuit is

$$\mathbf{V} = \mathbf{K}_a(i\omega)(\mathbf{E} + \mathbf{V}_{fb}) = \mathbf{K}_a(i\omega)[\mathbf{E} + \mathbf{K}_{fb}(i\omega)\mathbf{V}]$$

Solving this equation for $\mathbf{V}$, we obtain

$$\mathbf{V} = \frac{\mathbf{K}_a(i\omega)}{1 - \mathbf{K}_a(i\omega)\,\mathbf{K}_{fb}(i\omega)}\,\mathbf{E}$$

whence it follows that

$$\mathbf{K}_0(i\omega) = \frac{\mathbf{V}}{\mathbf{E}} = \frac{\mathbf{K}_a(i\omega)}{1 - \mathbf{K}_a(i\omega)\,\mathbf{K}_{fb}(i\omega)} \tag{5.78}$$

This expression is the main one for a system with feedback, $\mathbf{K}_0(i\omega)$ is sometimes called *total transfer function* or *closed-system transfer function*. On the other hand, the product $\mathbf{K}_a(i\omega)\mathbf{K}_{fb}(i\omega)$, which has the meaning of the transfer function of the cascade-connected two-port networks $\mathbf{K}_a(i\omega)$ and $\mathbf{K}_{fb}(i\omega)$, is called *open-system transfer function*.

Replacing $i\omega$ by p, we obtain the closed-system transfer function in the operator form:

$$\mathbf{K}_0(p) = \mathbf{K}_a(p)/[1 - \mathbf{K}_a(p)\,\mathbf{K}_{fb}(p)] \tag{5.79}$$

Comparison between $\mathbf{K}_0(i\omega)$ and $\mathbf{K}_a(i\omega)$ makes it possible to determine the sense of feedback in the general case where these functions are complex. If at some frequency the inequality $K_0(\omega) < K_a(\omega)$ holds, i.e., if the introduction of feedback reduces amplification, feedback at the given frequency is negative, otherwise it is positive.

At $\mathbf{K}_a(i\omega)\,\mathbf{K}_{fb}(i\omega) = 1$ the amplification $\mathbf{K}_0(i\omega)$ becomes infinitely high. This means that the circuit becomes unstable, and other technique should be used for its investigation, since expressions (5.67) through (5.78) are applicable only under steady-state conditions.

The case of unstable quiescent conditions is discussed in connections with the properties of self-oscillating systems (see Chap. 9). The present chapter deals with stable circuits only. The conditions of stability will be defined in Sec. 5.10 after presenting the basics of the theory of stability of linear circuits with feedback.

5.9. USE OF NEGATIVE FEEDBACK FOR IMPROVING AMPLIFIER CHARACTERISTICS

Let us consider the effect of feedback on the following amplifier parameters:

(a) stability of the amplification factor;

(b) nonlinear signal distortion due to the curvature of the current-voltage characteristic of amplifying devices;

(c) uniformity of the frequency characteristic within a given frequency band.

Assume that in a linear circuit acted upon by a harmonic e.m.f. and provided with a feedback loop one of the parameters $-\mathbf{K}_a(i\omega)$ or $\mathbf{K}_{fb}(i\omega)$ has changed either in modulus or in argument. Such a change may be caused by supply voltage fluctuations, changes in the ambient temperature, mechanical vibrations affecting the electrical parameters of the system, etc. Let us examine the effect of feedback on the relative change in the output signal. First, let us consider the case of instability of the amplifier circuit itself. To simplify the analysis, assume that prior to a change in the operating conditions, the transmission coefficients $\mathbf{K}_a(i\omega)$ and $\mathbf{K}_{fb}(i\omega)$ were purely real numbers K_a and K_{fb}, so that the closed-system transmission coefficient was defined by the expression

$$K_0 = K_a/(1 - K_aK_{fb}) \tag{5.80}$$

Assume the change due to instability consisted in that the coefficient K_a changed by a small value ΔK_a. Without feedback, this would result in a relative change of $\Delta K_a/K_a$ in the amplitude of the output voltage (the amplitude of the input e.m.f. is considered to be constant).

To determine the relative change in the amplitude in the presence of feedback, let us differentiate expression (5.80) with respect to K_a:

$$\frac{dK_0}{dK_a} = \frac{1}{(1-K_aK_{fb})^2} = \frac{K_a}{(1-K_aK_{fb})} \frac{1}{(1-K_aK_{fb})} \frac{1}{K_a}$$

whence

$$\frac{dK_0}{K_0} = \frac{1}{1-K_aK_{fb}} \frac{dK_a}{K_a} \tag{5.81}$$

From this expression it is clear that the relative change in the output voltage (i.e., the quantity dK_0/K_0 in the presence of feedback may considerably differ from that which would take place in the absence of feedback.

If feedback is *negative* ($K_aK_{fb} < 0$), the system *instability is reduced*:

$$\frac{dK_0}{K_0} = \frac{1}{1+|K_aK_{fb}|} \frac{dK_a}{K_a}$$

With *positive* feedback, the *instability is increased*:

$$\frac{dK_0}{K_0} = \frac{1}{1-|K_aK_{fb}|} \frac{dK_a}{K_a}$$

From this it follows that amplification stability can be increased by means of *negative* feedback. This method is widely used in modern radioelectronics. Depending on the requirements for the system stability, the magnitude $|K_aK_{fb}|$ may be as high as 100 and even higher. Naturally, in this case the amplification K_0 of the circuit is reduced by a factor of $(1 + |K_aK_{fb}|)$. This reduction can be compensated for by increasing K_a (for example, by increasing the number of stages in the circuit with feedback).

Now let us consider the instability of the feedback circuit. For this purpose, let us differentiate expression (5.80) with respect to K_{fb}:

$$\frac{dK_0}{dK_{fb}} = -\frac{K_a(-K_a)}{(1-K_aK_{fb})^2} = \frac{K_a}{1-K_aK_{fb}} K_0$$

whence

$$\frac{dK_0}{K_0} = \frac{K_aK_{fb}}{1-K_aK_{fb}} \frac{dK_{fb}}{K_{fb}}$$

In the case of negative feedback, where $|K_aK_{fb}| \gg 1$, we get

$$\frac{dK_0}{K_0} \approx -\frac{dK_{fb}}{K_{fb}}$$

From this relation it is clear that the influence of the *instability of the feedback circuit K_{fb} itself is not reduced by feedback*: the relative

instability of a closed circuit with negative feedback where $|K_a K_{fb}| \gg 1$ is equal to the relative instability of the quantity K_{fb}.

From this it follows that when using negative feedback, special attention should be given to the stability of the feedback circuit K_{fb}, the stability of the modulus and the argument (i.e., the phase characteristic) of the transfer function of the circuit being equally important. In practice, the feedback circuit K_{fb}, which usually is a simple passive network, can be made adequately stable with much greater ease than the amplifier circuit K_a, for it is exactly this circuit that contains the active and load elements responsible for the main causes of instability.

Now let us find out the effect of negative feedback on the nonlinear distortion due to the curvature of the characteristics of the active

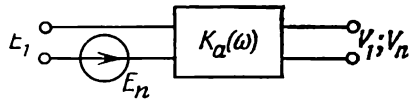


Fig. 5.23. Accounting for nonlinear distortion in an amplifier by means of an equivalent harmonic oscillator

elements in the main amplifier circuit. When the amplifier is excited by a harmonic voltage, this distortion manifests itself in the form of higher harmonics being present in the amplified signal. Assume the system has no feedback, and when it is excited by an e.m.f. E_1, the amplitude of the voltage of the main frequency at the output is equal to V_1, while that of one of the harmonics, V_n. An amplifier featuring nonlinear distortion may be represented in the form of an ideal linear amplifier whose input is acted upon by a "harmonics generator" (Fig. 5.23).

In this case the ratios E_n/E_1 and V_n/V_1 are equal, because the amplification factor $K_a(\omega)$ is considered to be the same for both the main frequency and the frequency of the nth harmonic. Thus, the amplitude of the e.m.f. of the equivalent generator E_n must be equal to V_n/K_a.

When negative feedback is used, the former amplitude V_1 at the output can be obtained by increasing the input e.m.f. E_1 by a factor of $(1 + |K_a K_{fb}|)$ as it follows from formula (5.80). This is shown in Fig. 5.24 in the form of an additional amplifier with an amplification factor $(1 + |K_a K_{fb}|)$. However, it should be remembered that the voltage of the main frequency acting directly across the terminals *3* and *3'* remains the same as in the circuit without negative feedback shown in Fig. 5.23. Indeed, the voltage in question is the difference between the e.m.f. $E_2 = E_1(1 + |K_a K_{fb}|)$

acting across the terminals *2* and *2'* (Fig. 5.24) and the feedback voltage $|K_{fb}|V_1$, i.e.,

$$E_3 = E_2 - K_{fb}V_1 = E_1(1 + |K_a||K_{fb}|) - |K_{fb}|V_1$$

But $E_1|K_a|$ is none other than V_1 (see Fig. 5.23), hence,

$$E_3 = E_1 + E_1|K_a||K_{fb}| - |K_{fb}|V_1 = E_1$$

The voltage of the nth harmonic at the amplifier input, with due regard for the feedback voltage $-V_nK_{fb}$ is equal to the difference $E_n - |K_{fb}|V_n$, while at the amplifier output

$$V_n = |K_a|(E_n - |K_{fb}|V_n)$$

whence

$$V_n = |K_a|E_n/(1 + |K_aK_{fb}|)$$

Thus, the ratio

$$\frac{V_n}{V_1} = \frac{|K_a|E_n}{(1+|K_aK_{fb}|)|K_a|E_1} = \frac{E_n/E_1}{1+|K_aK_{fb}|} \qquad (5.82)$$

is $(1 + |K_aK_{fb}|)$ times lower than in the absence of feedback. It is true, however, that this improvement is achieved by raising

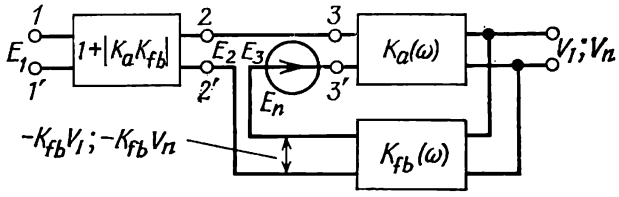


Fig. 5.24. Exlpaining the effect of reduction of the level of extraneous harmonics in an ampifier with negative feedback

the voltage applied across the terminals *2* and *2'* (Fig. 5.24) by a factor of $(1 + |K_aK_{fb}|)$.

The relative attenuation of the voltage of higher harmonics may also be explained in the following manner: the introduction of negative feedback results in a decrease of the gain by a factor of $(1 + |K_aK_{fb}|)$ both for the useful signal and for harmonics, however, this gain reduction is compensated for [by increasing the input voltage $(1 + |K_aK_{fb}|)$ times] for the useful signal only.

The above reasoning applies to all the harmonics of the voltage to be amplified.

Apart from reducing nonlinear distortion, the use of negative feedback also makes it possible (under certain conditions) to reduce ripple noises.

Thus, all spurious oscillations arising in the *amplifier proper* as a result of the nonlinearity of amplifying devices and imperfection of power units are attenuated by a factor of $(1 + | K_a K_{fb} |)$ through the use of negative feedback.

If an amplifier consists of several stages, it is good practice to apply negative feedback to all the stages as, for example, shown

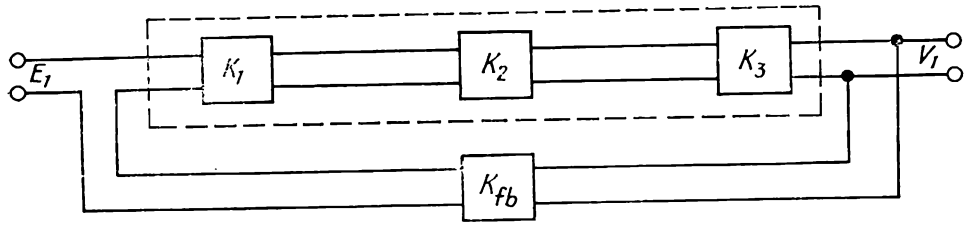


Fig. 5.25. Multistage amplifier with negative feedback

in Fig. 5.25. In this case, however, the stability of the amplifier may be disturbed, because of an increase in the total phase shift in the loop, particularly, where the amplifier uses transformers having stray inductance.

Where it is possible to build a multistage amplifier without transformers and with small parasitic capacitances, a circuit similar to that of Fig. 5.25 can be realized. Such conditions, for example, may be met with in transistor audio-frequency amplifiers.

In conclusion, let us consider the influence of negative feedback on the frequency characteristic of the amplifier. From expression (5.78) it directly follows that when $| \mathbf{K}_a (i\omega) \mathbf{K}_{fb} (i\omega) \gg | 1$

$$K_0 (\omega) \approx 1/K_{fb} (\omega) \tag{5.83}$$

If $K_{fb} (\omega)$ is kept constant within a specified frequency range, $K_0 (\omega)$ will also be constant. Thus the problem is reduced to levelling out the frequency characteristic of a passive two-port feedback circuit, and this is much easier than to eliminate the nonuniformity of the frequency characteristic of the amplifier $K_a (\omega)$.

Where $K_a K_{fb}$ amounts to a few unities, the ultimate relation (5.83) cannot be reached, however, the frequency characteristic $K_0 (\omega)$ becomes much more uniform than $K_a (\omega)$. This is illustrated in Fig. 5.26.

The dashed line in Fig. 5.26*a* shows the frequency characteristic of the aperiodic amplifier discussed in Sec. 5.6 (Fig. 5.14*b*). With the application of negative feedback with a real coefficient K_{fb}, the transfer function of the amplifier, in accordance with (5.78) and (5.44), will be

$$\mathbf{K}_0 (i\omega) = \frac{\mathbf{K}_a (i\omega)}{1 - \mathbf{K}_{fb} (i\omega) \mathbf{K}_a (i\omega)} = - \frac{K_{a\,\max}}{(1 + | K_{fb} K_{a\,\max} |) + i\omega C_0/(G_i + G)} \tag{5.84}$$

The modulus of this function, i.e., the amplitude-frequency characteristic,

$$K_0(\omega) = \frac{K_{a\,\max}}{\sqrt{(1+|K_{fb}K_{a\,\max}|)^2+\omega^2[C_0/(G_i+G)]^2}}$$

is shown in Fig. 5.26*a* by a continuous line. The characteristic is plotted on the basis of the following data: $K_{a\,\max} = 50$, $|K_{fb}| = 0.05$.

Thus,

$$K_0(\omega) = \frac{50}{\sqrt{(1+2.5)^2+[\omega C_0/(G_i+G)]^2}}$$

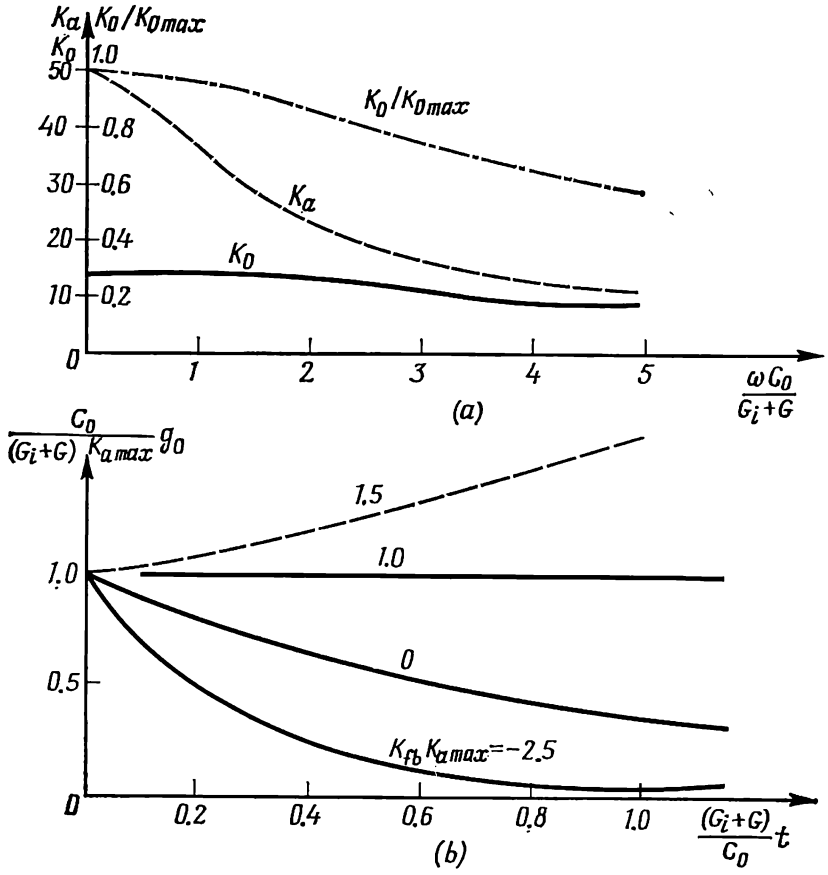


Fig. 5.26. (*a*) Frequency characteristics and (*b*) impulse responses of an amplifier with negative feedback

As one would expect, the curve $K_0(\omega)$ is located lower than the curve $K_a(\omega)$ (at all frequencies). This is a result of a voltage

from the amplifier output being applied to the input in *antiphase* with the input voltage. At frequencies close to zero, we have

$$K_{0\max} = \frac{K_{a\max}}{1+|K_{fb}K_{a\max}|} = \frac{50}{1+2.5} = \frac{1}{3.5}K_{a\max}$$

i.e., the gain is reduced 3.5 times.

However, the characteristic $K_0(\omega)$ is much more uniform than $K_a(\omega)$. This is clear from the normalized frequency characteristic, i.e., the function $K_0(\omega)/K_{0\max}$ (the dot-and-dash line in Fig. 5.26*a*). Thus, the use of negative feedback to stabilize amplification factor and reduce nonlinear distortion levels out at the same time the amplitude-frequency characteristic of the amplifier.

It should be noted that the required passband can be obtained without negative feedback, by correspondingly reducing the load resistance R. However, in this case the other amplifier parameters—stability and linearity—would be deteriorated.

The impulse response of the amplifier changes in accordance with the new characteristic $\mathbf{K}_0(i\omega)$. Indeed, writing expression (5.84) in a form similar to (5.43)

$$\mathbf{K}_0(i\omega) = -\frac{\dfrac{K_{a\max}}{1+|K_{fb}K_{a\max}|}}{1+\dfrac{i\omega C_0}{G_i+G}\dfrac{1}{1+|K_{fb}K_{a\max}|}} = -\frac{\dfrac{K_{a\max}}{1+|K_{fb}K_{a\max}|}}{1+i\omega\tau_{eq}}$$

we see that feedback results in a change in the equivalent time constant: instead of $C_0/(G_i+G)$ we obtain

$$\tau_{eq} = \frac{C_0}{G_i+G}\frac{1}{1+|K_{fb}K_{a\max}|}$$

At $|K_{fb}K_{a\max}| = 2.5$

$$\tau_{eq} = \frac{1}{3.5}\frac{C_0}{G_i+G}$$

It should be noted that the maximum gain (at $\omega = 0$) is reduced proportionally. Thus, if in the absence of feedback the impulse response of the amplifier in question is written in the form

$$g(t) = -\frac{K_{a\max}}{C_0/(G_i+G)}\exp\left[-\frac{(G_i+G)t}{C_0}\right]$$

with negative feedback it is

$$g_0(t) = \frac{K_{a\max}}{C_0/(G_i+G)}\exp\left[-\frac{(1+|K_{fb}K_{a\max}|)t}{C_0/(G_i+G)}\right]$$

The normalized impulse response $g_0(t)$ for several values of the parameter $|K_{fb}K_{a\max}|$ is shown in Fig. 5.26*b*.

As it might be expected, the use of negative feedback ($K_{fb}K_{a\,max} < 0$) that widens the passband of the circuit, results in a faster decrease in the impulse response. With positive feedback ($K_{fb}K_{a\,max} > 0$), the decrease in $g_0(t)$ is retarded. The dashed line in Fig. 5.26*b* shows the impulse response at $K_{fb}K_{a\,max} > 1$, which corresponds to unstable conditions (see Sec. 5.10).

5.10. STABILITY OF LINEAR ACTIVE CIRCUITS WITH FEEDBACK. ALGEBRAIC STABILITY CRITERION

A real circuit with feedback always contains energy-storing reactive elements. Even an amplifier built around resistors is not free from such elements (parasitic capacitances of the circuit and amplifying devices, inductances of the wires, etc.). The reactive elements produce additional phase shifts. If at some frequency these shifts make up a supplementary angle of 180°, feedback transforms from negative to positive, and conditions are created for parasitic oscillation to occur.

In many cases this circumstance considerably limits the use of feedback since at high values of $|K_aK_{fb}|$ spurious oscillation can be eliminated only by means of special phase compensators and others devices which reduce the slope of the phase-shift characteristic of the feedback loop. But it often turns out that the insertion of new elements into the system only shifts the frequency of spurious oscillations into the region of either very low or very high frequencies.

Thus, the use of feedback is intimately connected with the problem of stability of radio circuits.

Methods of determining the stability of a circuit are of paramount importance for designing the circuit and selecting its parameters. At present, several criteria are known, which differ in form, rather than in essence. Most of them are based on the stability criterion for the solutions of the differential equation describing the circuit under analysis.

Let the linear homogeneous equation of a circuit with lumped (and constant) parameters be given in the form

$$b_0\frac{d^nx}{dt^n}+b_1\frac{d^{n-1}x}{dt^{n-1}}+b_2\frac{d^{n-2}x}{dt^{n-2}}+\ldots+b_{n-1}\frac{dx}{dt}+b_nx=0 \qquad (5.85)$$

where x is current, voltage, etc., while the constant coefficients $b_0, b_1, b_2, b_3, \ldots, b_n$ are real numbers depending on the parameters of the circuit.

As we know, the solution of equation (5.85) has the form

$$x=\sum_{i=1}^{n}A_ie^{p_it}$$

where A_i are constants and p_i are the roots of the characteristic equation

$$b_0p^n + b_1p^{n-1} + b_2p^{n-2} + \ldots + b_{n-1}p + b_n = 0 \qquad (5.86)$$

The condition for the stability of a circuit in a state of equilibrium is that it must return to its initial state when the action of external disturbances has ceased. For this to happen, it is necessary that the transient (free) currents and voltages (arising in the system when its state of rest has been disturbed) be decaying. This, in turn, means that the roots $p_1, p_2, \ldots, p_n$ of equation (5.86) must be either real negative numbers or complex numbers with negative real parts. From these simple physical concepts stems the following fundamental stability criterion for any linear system: *a system is stable if the real parts of all the roots of the characteristic equation are negative**. It should be noted that the left-hand side of the characteristic equation (5.86) is none other than the denominator of the circuit transfer function written in the form

$$\mathbf{K}(p) = \frac{a_0p^n + a_1p^{n-1} + \ldots + a_{n-1}p + a_n}{b_0p^m + b_1p^{m-1} + \ldots + b_{m-1}p + b_m}, \quad n \leqslant m \qquad (5.87)$$

Thus, *the roots of the characteristic equation of a circuit are the poles of the transfer function* $\mathbf{K}(p)$ *of this circuit.*

From this it follows that the above-defined condition for the negativity of the real parts of the roots is equipollent to the following

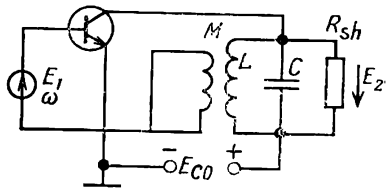


Fig. 5.27. Determining the stability of an amplifier with feedback

statement: *a circuit is stable if its transfer function* $\mathbf{K}(p)$ *has no poles on the right-hand half-plane of the complex variable* p.

This proposition, which is well known in the theory of circuits, may also be applied to the transfer function $\mathbf{K}_0(p)$ of a circuit with feedback.

Let us illustrate this with an example of a resonance amplifier with feedback (Fig. 5.27).

Let us determine the transfer function (as to voltage) of the amplifier by formula (5.60), assuming that $G_i \ll Y_l$:

$$\mathbf{K}_a(i\omega) = -\frac{S}{(1/i\omega L + i\omega C + 1/R_{sh})}$$

* This fundamental proposition was proved by A. M. Lyapunov who in the nineties of the last century laid down the foundations of the stability theory. The problem of stability of a system in a state of equilibrium is a particular case of Lyapunov's general stability theory.

The transfer function of the feedback circuit is $K_{fb}(i\omega) = \pm M/L$, where M is the coefficient of mutual inductance.
Replacing $i\omega$ by p, we reduce $\mathbf{K}_a(i\omega)$ to the form

$$\mathbf{K}_a(p) \approx -\frac{Sp}{C(p^2+p/R_{sh}C+1/LC)} = -\frac{Sp}{C(p^2+2\alpha_c p+\omega_r^2)}$$

where $\alpha_c = 1/2\ R_{sh}C$ and $\omega_r^2 = 1/LC$. Then the transfer function of the amplifier with feedback will be

$$\mathbf{K}_0(p) = \frac{\mathbf{K}_a(p)}{1-\mathbf{K}_a(p)\mathbf{K}_{fb}} = -\frac{S}{C}\frac{p}{p^2+(2\alpha_c+\mathbf{K}_{fb}S/C)\,p+\omega_r^2}$$

Let us find the roots of the equation $p^2 + (2\alpha_c + \mathbf{K}_{fb}S/C)\,p + \omega_r^2 = 0$:

$$p_{1,2} = -\alpha_c - \frac{\mathbf{K}_{fb}S}{2C} \pm i\sqrt{\omega_r^2 - \left(\alpha_c + \frac{\mathbf{K}_{fb}S}{2C}\right)^2}$$

Let us consider two possible cases —negative and positive feedback.
To effect negative feedback, the product $\mathbf{K}_a\mathbf{K}_{fb}$ must be negative.
Since at $\omega = \omega_r$, i.e., at resonance, $\mathbf{K}_a(i\omega)$ is negative, the coefficient K_{fb} must be positive: $K_{fb} = +M/L$. In this case the real parts of both roots p_1 and p_2

$$\mathrm{Re}\,(p_{1,2}) = -(\alpha_c + MS/2LC)$$

are negative for any value of M.
With positive feedback, $K_{fb} = -M/L$;

$$\mathrm{Re}\,(p_{1,2}) = -\left(\alpha_c - \frac{MS}{2LC}\right) \left.\begin{array}{l} <0 \text{ for } MS/2LC < \alpha_c \\ >0 \text{ for } MS/2LC > \alpha_c \end{array}\right\} \quad (5.88)$$

Thus, with negative feedback, the circuit in question is stable at any value of M, while with positive feedback, it is stable only when the condition

$$|K_{fb}| = \frac{M}{L} < \frac{2C\alpha_c}{S} = \frac{1}{SR_{sh}} = \frac{1}{K_{a\max}}$$

is satisfied, $K_{a\max} \approx SR_{sh} = S/G_{sh}$ being the amplification factor at the resonance frequency [see (5.68)].
When the circuit is described by a higher-degree differential equation, the study of the roots of the characteristic equation, necessary for determining whether the system is stable, is a complicated problem.
It has been found that this problem can be solved by analyzing the relations between the coefficients of the equation without determining the roots themselves. This can be done using the Hurwitz

theorem which reads as follows: for the real parts of all the roots of the equation

$$b_0x^m + b_1x^{m-1} + b_2x^{m-2} + \ldots + b_{m-1}x + b_m = 0$$

with real coefficients and $b_0 > 0$ to be negative, it is necessary and sufficient that all the determinants set up of the coefficients $b_0, b_1, b_2, \ldots, b_n$ according to the form

$$\Delta_1 = b_1; \quad \Delta_2 = \begin{vmatrix} b_1 & b_3 \\ b_0 & b_2 \end{vmatrix},$$

$$\Delta_3 = \begin{vmatrix} b_1 & b_3 & b_5 \\ b_0 & b_2 & b_4 \\ 0 & b_1 & b_3 \end{vmatrix}, \quad \Delta_4 = \begin{vmatrix} b_1 & b_3 & b_5 & b_7 \\ b_0 & b_2 & b_4 & b_6 \\ 0 & b_1 & b_3 & b_5 \\ 0 & b_0 & b_2 & b_4 \end{vmatrix},$$

$$\Delta_5 = \begin{vmatrix} b_1 & b_3 & b_5 & b_7 & b_9 \\ b_0 & b_2 & b_4 & b_6 & b_8 \\ 0 & b_1 & b_3 & b_5 & b_7 \\ 0 & b_0 & b_2 & b_4 & b_6 \\ 0 & 0 & b_1 & b_3 & b_5 \end{vmatrix}, \text{ etc.}$$

be positive.

The algebraic stability criterion is often referred to as the *Routh-Hurwitz criterion*. When setting up the determinants according to the above scheme, the coefficients with an index exceeding the degree of the characteristic equation are replaced by zeros.

For example, in the case of a fourth-degree equation, the determinants are as follows

$$\Delta_1 = b_1, \quad \Delta_2 = \begin{vmatrix} b_1 & b_3 \\ b_0 & b_2 \end{vmatrix}$$

$$\Delta_3 = \begin{vmatrix} b_1 & b_3 & 0 \\ b_0 & b_2 & b_4 \\ 0 & b_1 & b_3 \end{vmatrix}, \quad \Delta_4 = \begin{vmatrix} b_1 & b_3 & 0 & 0 \\ b_0 & b_2 & b_4 & 0 \\ 0 & b_1 & b_3 & 0 \\ 0 & b_0 & b_2 & b_4 \end{vmatrix}$$

It is evident that all successive determinants are the principal diagonal minors of the determinant Δ_m. Since the last column of the determinant Δ_m contains only one nonzero element b_m located on the principal diagonal, the following equation is satisfied:

$$\Delta_m = b_m\Delta_{m-1}$$

From this it follows that, in accordance with the Hurwitz theorem, the stability conditions may be stated in terms of the following inequalities:

$$\Delta_1 > 0,\ \Delta_2 > 0,\ \ldots,\ \Delta_{m-1} > 0,\ b_m > 0$$

Thus, for example, for a second-degree characteristic equation we have

$$\Delta_1 = b_1 > 0,\ b_2 > 0 \tag{5.89}$$

For a third-degree equation

$$\Delta_1 = b_1 > 0$$

$$\Delta_2 = \begin{vmatrix} b_1 & b_3 \\ b_0 & b_2 \end{vmatrix} = b_1 b_2 - b_3 b_0 > 0,\quad b_3 > 0 \tag{5.90}$$

i.e., $b_1 > 0$, $b_1 b_2 > b_3 b_0$, $b_3 > 0$. Since b_0, b_1 and b_3 are positive, $b_2 > 0$.

For a fourth-degree equation

$$\text{I. } \Delta_1 = b_1 > 0;$$

$$\text{II. } \Delta_2 = b_1 b_2 - b_3 b_0 > 0;$$

$$\text{III. } \Delta_3 = b_3 (b_1 b_2 - b_3 b_0) - b_1^2 b_4 > 0;$$

$$\text{IV. } b_4 > 0.$$

From condition III, and subject to conditions IV and I, we have the following inequality:

$$b_3 (b_1 b_2 - b_0 b_3) > b_1^2 b_4 > 0$$

Therefore, the second condition can be replaced by the condition $b_3 > 0$. Thus, for a fourth-degree equation, we have the following stability conditions:

$$b_1 > 0,\ b_3 > 0,\ b_3 (b_1 b_2 - b_0 b_3) - b_1^2 b_4 > 0,\ b_4 > 0 \tag{5.91}$$

Let us illustrate the use of the Routh-Hurwitz criterion with a simple example of the above-discussed resonance amplifier with feedback (Fig. 5.27). The characteristic equation of this circuit at $K_{fb} = M/L$ (negative feedback) is

$$p^2 + \left(2\alpha_c + \frac{M}{L}\frac{S}{C}\right) p + \omega_r^2 = 0$$

The stability conditions (5.89) set up for a second-degree equation in the given case have the form

$$\Delta_1 = b_1 = 2\alpha_c + \frac{M}{L}\frac{S}{C} > 0,\quad b_2 = \omega_r^2 > 0$$

The first condition is satisfied at any value of M, while the second one is independent of M.

With positive feedback ($K_{fb} = -M/L$), the circuit is stable, provided that the condition $2\alpha_c - (M/L)(S/C) > 0$, which coincides with (5.88), is satisfied.

The Routh-Hurwitz criterion is particularly useful for checking the stability of a circuit with specified parameters (i.e., the coefficients of the differential equation describing the circuit). However, its experimental utilization is not convenient since in this case we do not know the coefficients of the equation but only know the open-circuit transfer function $\mathbf{K}_a(p)\,\mathbf{K}_{fb}(p)$. In addition, the Routh-Hurwitz criterion does not indicate clearly how to make stable an unstable circuit.

5.11. FREQUENCY STABILITY CRITERIA

The requirement that the transfer function

$$\mathbf{K}_0(p) = \mathbf{K}_a(p)/[1 - \mathbf{K}_a(p)\,\mathbf{K}_{fb}(p)] \qquad (5.92)$$

should have no poles on the right-hand half-plane $p = \sigma + i\omega$, i.e., in the region limited by a semicircle of infinitely large radius R and by the axis $i\omega$ (Fig. 5.28*a*) is equivalent* to the condition

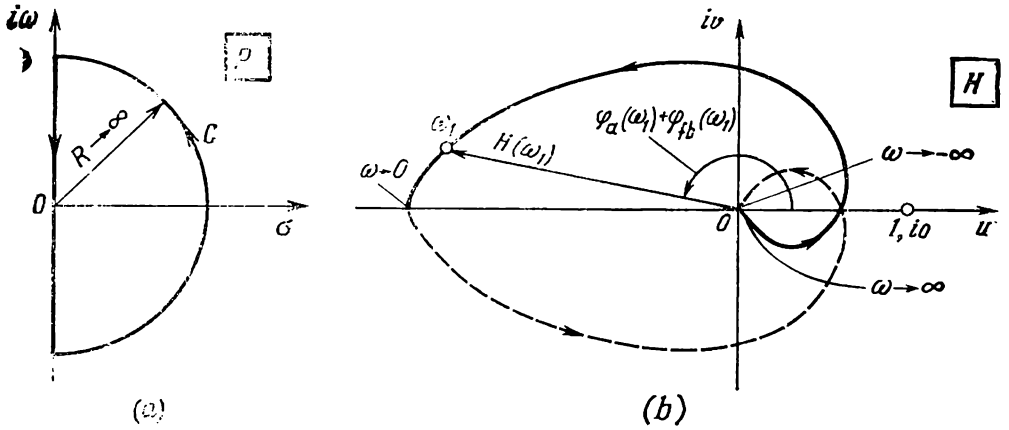


Fig. 5.28. (a) Closed contour on the p-plane and (*b*) hodograph of the function $H(i\omega)$ on the plane $u + iv$

that the denominator of expression (5.92) should not have zeros in this region or, which is the same, the function

$$\mathbf{H}(p) = \mathbf{K}_a(p)\,\mathbf{K}_{fb}(p) \qquad (5.93)$$

should not turn into unity at any point of the right-hand half-plane p. But $\mathbf{H}(p)$ is the transfer function of the open-circuit feedback loop, i.e., the ratio of the voltage across the terminals *2-2*

* It is supposed that the amplifier circuit proper is stable, i.e. $\mathbf{K}_a(p)$ has no poles in the right-hand half-plane p.

to that across the terminals *1-1* when the loop is open-circuited, as shown in Fig. 5.29. Therefore, we may judge of the stability of a system with feedback by the characteristics of the open-circuited system.

For further analysis, it is expedient to transform from the plane $p = \sigma + i\omega$ to the plane $\mathbf{H}(p) = u + iv$ (Fig. 5.28*b*). Each point p on the plane σ, $i\omega$ corresponds to a definite value of $\mathbf{H}$ on the plane u, iv. Any closed contour on the plane p can be transformed to some contour (also closed) on the plane $\mathbf{H}$ by means of expression (5.93).

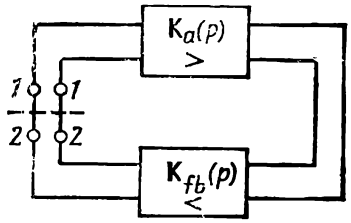


Fig. 5.29. Determining the transfer function of the open-circuited amplifier — two-port feedback network system

If the initial contour on the plane p is given in the form shown in Fig. 5.28*a*, the corresponding contour on the plane $\mathbf{H}$ is called the *hodograph* of the function $\mathbf{H}$.

The contour C shown in Fig. 5.28*a* can be divided into two sections: (1) a straight line $i\omega$, with ω varying from ∞ to $-\infty$, and (2) semicircle of infinitely large radius R.

In the first section, where $\sigma = 0$, $p = i\omega$, the function $\mathbf{H}(p)$ transforms to a function $\mathbf{H}(i\omega)$. In accordance with expression (5.93), this section is transformed on the plane $\mathbf{H}$ to a line defined by the following relation:

$$\mathbf{H}(i\omega) = \mathbf{K}_a(i\omega)\,\mathbf{K}_{fb}(i\omega) = K_a(\omega)\,K_{fb}(\omega)\,e^{i(\varphi_a+\varphi_{fb})} = u(\omega) + iv(\omega) \tag{5.94}$$

whence

$$\begin{aligned} u(\omega) &= K_a(\omega)K_{fb}(\omega)\cos(\varphi_a + \varphi_{fb}) \\ v(\omega) &= K_a(\omega)\,K_{fb}(\omega)\sin(\varphi_a + \varphi_{fb}) \end{aligned} \tag{5.95}$$

In these expressions, φ_a and φ_{fb} are the arguments of the transfer functions of the two-port networks $\mathbf{K}_a(i\omega)$ and $\mathbf{K}_{fb}(i\omega)$, respectively.

In the second section of the contour C (Fig. 5.28*a*), as $R \to \infty$ the function* $\mathbf{H}(p) \to 0$. This stems from the general expression (5.87) which may be written in the form

$$\mathbf{K}(p) = B\,\frac{(p-p_{01})(p-p_{02})\ldots(p-p_{0n})}{(p-p_{p1})(p-p_{p2})\ldots(p-p_{pm})} \tag{5.96}$$

where B is a constant coefficient, and p_{0i} and p_{pj} are the zeros and poles of the function $\mathbf{K}(p)$, respectively, the poles being in the left-hand half-plane. With $|p| \to \infty$, the quantities p_{0i} and p_{pj} may

* Here we mean most widely used two-port networks whose transfer function has the exponent n of the numerator smaller than the exponent m of the denominator.

be disregarded and then function $\mathbf{K}(p)$ may be represented in the form $Bp^{(n-m)}$. This holds true for both the left- and right-hand half-planes. In a similar way, the function $\mathbf{H}(p)$, with $|p| \to \infty$, may be represented in the form

$$\mathbf{H}(p) = Ap^{(n-m)}$$

where n and m are the numbers of the zeros and poles of the function $\mathbf{H}(p)$.

With $n < m$ and $|p| \to \infty$, the modulus of the function $\mathbf{H}(p)$ on the semicircle $R \to \infty$ is zero. Thus, the semicircle of infinitely large radius R on the plane p is transformed to a point lying at the origin of coordinates on the plane $\mathbf{H}$, and to construct the hodograph $\mathbf{H}$ in the form of a closed contour, it is sufficient to know the behaviour of $\mathbf{H}(p)$ on the axis $i\omega$, i.e., to know the amplitude-frequency and phase-frequency characteristics of the circuit $\mathbf{K}_a(i\omega)\mathbf{K}_{fb}(i\omega)$.

The circuit of the hodograph $\mathbf{H}$ with frequency varying from ∞ to $-\infty$ corresponds to the circuit of the contour C in the positive direction (i.e., counterclockwise).

The entire right-hand half-plane p is transformed on the plane $\mathbf{H}$ to the interior of the hodograph. Therefore, if the hodograph *of the transfer function of the open-circuited system does not encompass the point 1, i0, then, with the closed feedback circuit the system is stable; otherwise it is unstable.*

This condition is known as the *Nyquist stability criterion.*

The diagram shown in Fig. 5.28*b* corresponds to a stable system. This is clear from the fact that the hodograph $\mathbf{H}$ does not encompass the point *1, i0*. The continuous line shows the part of the contour corresponding to positive frequencies $0 < \omega < \infty$, while the dashed curve shows the part corresponding to negative frequencies. Since the function $u(\omega)$ is even and $v(\omega)$, odd with respect to ω, both parts of the hodograph are symmetrical relative to the real axis.

It should also be noted that Figure 5.28*b* is constructed for the case where $\omega = 0$ and the transfer function $\mathbf{H}(i\omega)$ is nonzero (this is possible, for example, in d-c amplifiers that have no blocking capacitors).

In the case of complex circuits, the shape of the hodograph may sometimes be so involved that it is difficult to see whether the point *1, i0* is encircled by the hodograph. In such cases a method based on the Nyquist criterion and consisting in counting the number of intercepts made by the hodograph on the axis $u(\omega)$ in the region from 1 to ∞ proves to be useful. For a circuit to be stable, it is necessary that the hodograph either should not intersect this section of the $u(\omega)$-axis at all (see Fig. 5.28*b*) or should intersect it the same number of times in the positive and negative directions.

The Nyquist criterion is most widely used in radioelectronics, automation, and other allied fields. Its main advantage is that it allows one to manipulate conveniently the amplitude-frequency and phase-frequency characteristics of an open-circuit system. In some systems, like those containing lines, it is essentially the only acceptable method.

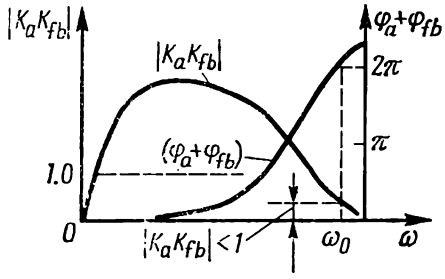


Fig. 5.30. Example of the amplitude-frequency and phase-frequency characteristics of a stable amplifier with feedback loop

With the Nyquist criterion, one can use ordinary amplitude-frequency and phase-frequency characteristics of an open-circuit system instead of the polar diagrams (hodographs) shown in Fig. 5.28.

Indeed, the length of the vector $\mathbf{H}(i\omega)$, as seen from expression (5.94), is none other than the modulus of the transfer factor of the open circuit K_aK_{fb}, i.e., the frequency characteristic of this circuit, while the argument φ_H (Fig. 5.30) equal to

$$\varphi_H = \arctan \frac{v(\omega)}{u(\omega)} = \varphi_a(\omega) + \varphi_{fb}(\omega) \qquad (5.97)$$

is the phase characteristic of the circuit K_aK_{fb}.

Having combined the amplitude and phase characteristics on a common graph, one can readily answer the question whether the system is stable.

If during the change of ω from 0 to ∞ the phase φ_H does not reach the value 2π, the closed circuit is stable at any value of K_aK_{fb}. On the other hand, if K_aK_{fb} is less than unity at any frequency, the circuit is stable no matter what its phase characteristic. The circuit is unstable if there are frequencies at which the following two conditions are satisfied simultaneously:

$$\varphi_a + \varphi_{fb} = n2\pi \ (n = \text{integer}), \ H = K_aK_{fb} \geqslant 1 \qquad (5.98)$$

In fact, these two conditions are necessary for turning into zero the denominator of expression (5.78) which defines the transfer function of the closed circuit.

An example of amplitude-frequency and phase-frequency characteristics of a stable circuit with feedback is shown in Fig. 5.30, and that of an unstable circuit is shown in Fig. 5.31. In the first case (Fig. 5.30), at a frequency ω_0 corresponding to $\varphi_a + \varphi_{fb} = 2\pi$ the modulus $H < 1$. In the second case (Fig. 5.31), ω_g is the spurious oscillation frequency. In Figs. 5.30 and 5.31 are plotted the magnitudes of $\varphi_a + \varphi_{fb}$. If the sense of real φ_a and φ_{fb} is taken into account, the slope of the phase characteristics will be negative.

When constructing these characteristics, account has been taken of the fact that when $\omega = 0$ and $\omega = \infty$ the value of $K_a K_{fb}$ is zero. With $\omega \to 0$, this is due to the influence of the series capacitors in the K_a or the K_{fb} circuit while with $\omega \to \infty$, this is due to the

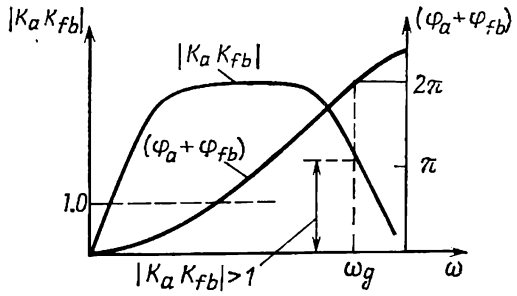


Fig. 5.31. The same for an unstable amplifier

effect of the shunt capacitances (interelectrode capacitances, wiring capacitances, etc.). The total phase change with the variation of ω from 0 to ∞ depends on the character and number of links in the amplifier and the feedback circuit.

Chapter 6

TRANSMISSION OF DETERMINISTIC OSCILLATIONS THROUGH LINEAR CIRCUITS WITH CONSTANT PARAMETERS

6.1. GENERAL

Radioelectronics deals with different types of signal and various, mostly time-lag, networks. When transmitting signals through such networks, there occur transient processes which affect the shape of the signals and, in the final analysis, the information carried by them. In Chap. 1, it was noted that the majority of radio devices are combinations of linear and nonlinear elements. This fact complicates the problem of rigorous analysis of transient processes in radio circuits, since the classical methods of investigation, based on the use of the superposition principle, are linear. That is why approximate methods for analyzing the action of signals on real devices have gained recognition in radio engineering practice. Here, linear circuits are isolated and studied separately from nonlinear elements and, when studying the transmission of signals through oscillatory circuits having good frequency selectivity, the analysis is simplified by assuming amplitude variations to be slow.

In spite of the above limitations, a great number of practical problems can be solved by linear methods. Such problems, first of all, arise when considering transmission of signals through linear amplifiers with simple- and quadratic-lag circuits. From the text below it will be evident that the weakly pronounced (in low-level signal operation) nonlinearity of amplifying devices (electron tubes, transistors, etc.) allows the use of linear methods to analyze the transmission of pulses and modulated waves through amplifier circuits. Even in the case of substantially nonlinear devices, linear analysis of their individual units often yields useful results.

Let us recall the basic methods employed when analyzing the transmission of signals through radio circuits.

In the case of simple circuits described by the first- or second-order differential equations, the problem can usually be solved with ease by the classical method of differential equations.

For complex circuits, the Fourier integral method and the closely associated operator method (Laplace transformations) prove much more convenient. Along with the spectral method, radioelectronics also frequently uses the superposition integral method whereby a signal is represented as a sum of pulses (or steps).

In addition to the exact methods, application find the already mentioned approximate methods adapted to specific circuits and

signals. This chapter presents the basic propositions of the theory of transmission of deterministic signals through linear circuits with constant parameters.

6.2. SPECTRAL METHOD

This method is based on the use of the transfer function $\mathbf{K}(i\omega)$ of a circuit, which was introduced in the preceding chapter (see Sec. 5.3).

If an arbitrarily-shaped signal in the form of an r.m.f. $e(t)$ acts at the input of a linear two-port network, then, by the spectral method, one should find the spectral density $\mathbf{E}(\omega)$ of the input signal. This operation can easily be done by means of expression (2.48). After that, multiplying $\mathbf{E}(\omega)$ by $\mathbf{K}(i\omega)$, one gets the spectral density of the signal at the output of the two-port network and finally, applying the inverse Fourier transformation [see expression (2.49)] to the product $\mathbf{E}(\omega)\,\mathbf{K}(i\omega)$, one finds the output signal in the form of a function of time.

Thus, if the input signal is written as the integral

$$e(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{E}(\omega)\,e^{i\omega t}\,d\omega \tag{6.1}$$

the output signal can be written in a similar form

$$v(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{E}(\omega)\,\mathbf{K}(i\omega)\,e^{i\omega t}\,d\omega \tag{6.2}$$

Comparison between expressions (6.2) and (6.1) shows that *the signal at the output of a linear circuit can be obtained by summation of the components of the input signal spectrum* $\mathbf{E}(\omega)$, *taken with weight* $\mathbf{K}(i\omega)$. In other words, the transfer function $\mathbf{K}(i\omega)$ is a *weighting function* that determines the relative contribution of the various components of the spectrum $\mathbf{E}(\omega)$ to the signal $v(t)$.

In Sec. 2.13 it was pointed out that the analysis of transient processes is much simplified when both the external forces (stimulus) and the transfer function of the circuit are represented as Laplace transforms. In this case, one may preserve the designation of the transfer function and change only the argument so that $\mathbf{K}(i\omega)$ is transformed to $\mathbf{K}(p)$, while the function $\mathbf{E}(\omega)$ is transformed to $L_e(p)$ (see Sec. 2.13). To simplify notation, the Laplace transform of the time function $e(t)$ will be designated $\overline{\mathbf{E}}(p)$. Then expression (6.2) will be reduced to the form [see Sec. 2.13]

$$v(t)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\overline{\mathbf{E}}(p)\,\mathbf{K}(p)\,e^{pt}\,dp \tag{6.3}$$

For $t > 0$, the closed integration contour formed by adding an arc of infinitely large radius in the left-hand half-plane (Fig. 6.1) encircles all the poles of the integrand, i.e., those of both $\bar{\mathrm{E}}(p)$ and $\mathrm{K}(p)$, owing to which the following relation holds:

$$v(t) = \frac{1}{2\pi i} \oint \bar{\mathrm{E}}(p)\,\mathrm{K}(p)\,e^{pt}\,dp = \sum \text{res}, \quad t > 0 \tag{6.4}$$

Here $\sum$ res is the sum of residues at the poles.

For $t < 0$, the integration contour lies in the right-hand half-plane and contains no poles, so the integral is zero.

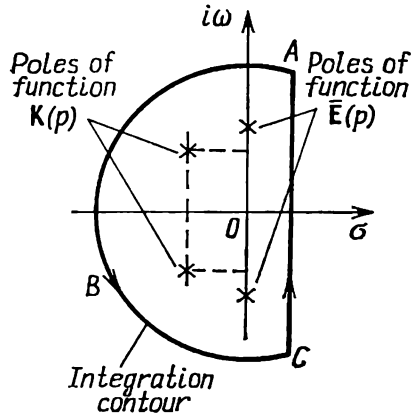


Fig. 6.1. Integration contour for $t > 0$

The arrangement (on the imaginary axis) of the poles of the function $\bar{\mathrm{E}}(p)$, shown in Fig. 6.1, corresponds to an e.m.f. of the form $e(t) = E_0 \cos \omega_0 t$, which exists for $t \geqslant 0$.

Thus, the calculation of the integral (6.4) is reduced to determining the residues at the poles of the integrand.

Let us represent the integrand of expression (6.4) in the form

$$\bar{\mathrm{E}}(p)\,\mathrm{K}(p)\,e^{pt} = \bar{\mathrm{V}}(p)\,e^{pt} = P(p)/Q(p) \tag{6.5}$$

Then the residue of the function $P(p)/Q(p)$ having a simple pole (of the first order) at the point p_1 is defined by the expression

$$\text{res}_1 = P(p_1) \Big/ \left[\frac{dQ(p)}{dp}\right]_{p=p_1} \tag{6.6}$$

If the function $P(p)/Q(p)$ has a pole of order m (m being a positive integer) at the point p_1, we have

$$\text{res}_1 = \frac{1}{(m-1)!} \frac{d^{m-1}}{dp^{m-1}} \left[\frac{P(p)}{Q(p)} (p - p_1)^m\right]_{p=p_1} \tag{6.7}$$

The method of application of contour integrals for determining some functions playing an important part in the theory of transient processes will be illustrated by examples later in the text.

6.3. SUPERPOSITION INTEGRAL METHOD

Instead of resolving a compound signal into its harmonic components (as in the spectral method), we can divide this signal into sufficiently short pulses (Fig. 6.2).

While the spectral method employs the transfer function $K(i\omega)$ of a circuit, the superposition integral method is based on the impulse response $g(t)$ introduced in Sec. 5.3.

Let it be required to find the output signal $s_{out}(t)$ of a circuit, given the signal $s(t)$ acting at the input of this circuit and its impulse response $g(t)$.

To explain the essence of the superposition integral method, let us divide an arbitrary signal $s(x)$ into elementary pulses (as shown in Fig. 6.2) and find the circuit response at the moment t to an elementary pulse (shaded in Fig. 6.2) acting at the input at a time x. Should the area of this pulse be equal to unity, it would be possible to regard the pulse as a delta function developed at the instant x. With the impulse response $g(x)$, the circuit reaction at the moment t would then apparently be $g(t-x)$. However, since the pulse area shaded in Fig. 6.2 is equal to $s(x)\,\Delta x$ (and not to unity), the response of the circuit at the moment t is equal to $s(x)\,\Delta x g(t-x)$.

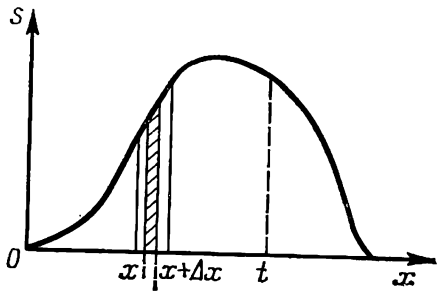


Fig. 6.2. Illustrating the superposition integral method

To find the total value of the output signal at the moment t, it is necessary to take the sum of the actions of all the pulses over the interval from $x = 0$ to $x = t$. With $\Delta x \to 0$, the summation is reduced to integration.

Then,

$$s_{out}(t) = \int_0^t s(x)\, g(t-x)\, dx \qquad (6.8)$$

In the general case, if the origin of the signal $s(x)$ does not coincide with the time origin x, the last expression may be written in the form

$$s_{out}(t) = \int_{-\infty}^{t} s(x)\, g(t-x)\, dx \qquad (6.9)$$

For real systems, the condition

$$g(t-x) = 0 \text{ for } t < x \qquad (6.10)$$

is always fulfilled, i.e., the function $g(t-x)$ must vanish when the argument becomes negative, since the response cannot come before the stimulus. In view of this, expression (6.8) can be replaced by

$$s_{out}(t) = \int_{-\infty}^{\infty} s(x)\, g(t-x)\, dx \qquad (6.11)$$

(one should remember that for $x > t$ the integrand vanishes).

Finally, let us write another expression obtained from (6.8) where x is replaced by $t - x$:

$$s_{out}(t) = \int_0^t s(t-x)\, g(x)\, dx \tag{6.12}$$

The integral on the right-hand side of equation (6.8) is known in mathematics as the convolution of the functions $s(t)$ and $g(t)$ (see Sec. 2.7). Thus, we arrive at the following important conclusion: *the output signal $s_{out}(t)$ of a linear circuit is the convolution of the input signal $s(t)$ with the impulse response $g(t)$ of the circuit.*

From expression (6.11) it is clear that at the instant t the signal $s_{out}(t)$ at the circuit output is the result of summation of all the previous instantaneous values of the input signal $s(t)$, taken with weight $g(t - x)$.

In Sec. 6.2 the transfer function $K(i\omega)$ served as a weighting function when summing the components of the spectrum. In our case, when summing the instantaneous values of the input signal $s(t)$, the *weighting function* is the *impulse response of the circuit* taken with the argument $t - x$, i.e., the function $g(t - x)$.

6.4. TRANSMISSION OF DISCRETE SIGNALS THROUGH AN APERIODIC AMPLIFIER

Discrete messages are usually carried by trains of pulses. When such trains are transmitted through time-lag circuits, the shape of the pulses undergoes changes leading to partial or complete loss of the information being transmitted. In this connection, the analysis of pulse-shape distortion is one of the typical problems facing a radio engineer.

From among pulses of numerous shapes, the rectangular pulse is of greatest interest. This is due to the simplicity of the pulse-forming networks and also the wide application of rectangular pulses in binary-code systems and many other radio engineering devices. When dealing with rectangular pulses, attention is usually concentrated on the transmission of the leading and trailing edges of a pulse. This is particularly important when the transmitted or received information is contained in the position of the leading (or trailing) edge of the pulse on the time axis (for example, in some radar systems).

Let us consider the passage of a rectangular pulse through the aperiodic amplifier studied in Sec. 5.6. First, let us consider the amplifier circuit including the separation circuit R_1, C_1 (see Fig. 5.15*a*). The transfer function of this amplifier is defined by expressions (5.50) through (5.52). Then, setting $C_1 \to \infty$, let us proceed to the circuit shown in Fig. 5.14*a*, that is typical of a transistor amplifier.

Assume that at the instant $t = 0$ the amplifier is excited by a rectangular e.m.f. pulse $e(t)$ of amplitude E and duration T (Fig. 6.3*a*). During the time interval from $t = 0$ to $t = T$ the voltage at the amplifier output may be regarded as being the result of application at $t = 0$ of a d-c e.m.f. $e_1(t) = E$. At the moment $t = T$ an additional e.m.f. $e_2(t) = -E$ is applied to cancel the e.m.f. $e_1(t)$, Fig. 6.3*b*. The superposition of the output voltages $v_1(t)$ and $v_2(t)$ due to the action of $e_1(t)$ and $e_2(t)$ forms a pulse at the amplifier output. Thus, the problem can be reduced to the study of the processes in the amplifier, that occur when a d-c e.m.f. is applied to the input.

In accordance with formula (2.108), when the e.m.f. $e_1(t) = E$ is applied at the instant $t = 0$, the Laplace transform of $e_1(t)$ is

$$\overline{E}_1(p) = \int_0^\infty e_1(t)\, e^{-pt}\, dp = E\frac{1}{p}$$

Then, according to formula (6.3), the output voltage is

$$v_1(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{1}{p} EK(p)\, e^{pt}\, dp$$

$$= -K_{\max} E \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{e^{pt}\, dp}{p^2\tau_0 + p + 1/\tau_1}$$

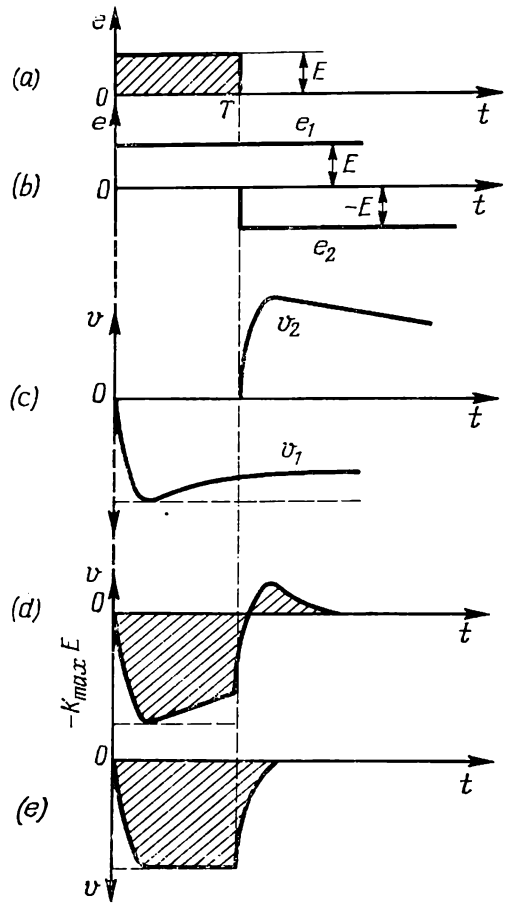


Fig. 6.3. Distortion of the shape of a pulse in a resistance coupled amplifier (*a*) input pulse; (*b*) representing the pulse as a sum of two steps; (*c*) deformation of the steps at the output; (*d*) resultant output pulse; (*e*) the pulse at the amplifier output with the separation circuit R_1C_1 removed

The poles of the integrand were determined in Sec. 5.6 [see expressions of (5.53)]:

$$p_1 = -1/\tau_1,\ p_2 = -1/\tau_0,\ \tau_0 \ll \tau_1$$

The residues are found by formula (6.6):

$$\text{res}_1 = \frac{e^{p_1 t}}{\left[\frac{dQ(p)}{dp}\right]_{p=p_1}} = \frac{e^{-t/\tau_1}}{2\tau_0\left(-\frac{1}{\tau_1}\right)+1} \approx e^{-t/\tau_1}$$

$$\text{res}_2 = \frac{e^{p_2 t}}{\left[\frac{dQ(p)}{dp}\right]_{p=p_2}} = \frac{e^{-t/\tau_0}}{2\tau_0\left(-\frac{1}{\tau_0}\right)+1} = -e^{-t/\tau_0}$$

Thus, we have

$$v_1(t) = -K_{\max}E\,[e^{-t/\tau_1} - e^{-t/\tau_0}] \qquad (6.13)$$

The graphs of $v_1(t)$ and $v_2(t) = -v_1(t-T)$ are shown in Fig. 6.3*c* and that of the resultant voltage $v(t) = v_1(t) + v_2(t)$ at the amplifier output, in Fig. 6.3*d*.

From formula (6.13) and Fig. 6.3*d* it is clear that while t is commensurable with τ_0, the first exponent in expression (6.13) is close to unity and the leading edge of the pulse is influenced mainly by the second exponent, but when t becomes commensurable with τ_1, the character of the function $v_1(t)$ is basically determined by the first exponent. The same applies to the function $v_2(t)$ when reading the time from the instant $t = T$. A square pulse with an amplitude $-K_{\max}E$ that would occur in an ideal amplifier with a uniform amplitude-frequency characteristic is shown in Fig. 6.3*d* by a dashed line.

The distortion of the shape of a real pulse manifests itself (a) in the finite slope of the pulse edges, and (b) in the decline of the pulse top.

The *greater the time constant* $\tau_0 = RC_0$ (and hence, the sharper the drop of the frequency characteristic curve in the *high-frequency region*), the more pronounced is the first of these factors, and conversely, the *smaller the time constant* τ_1 *of the separation circuit* R_1C_1 (and hence, the sharper the drop of the frequency characteristic curve in the *low-frequency region*), the more pronounced is the second factor.

The choice of the time constants τ_0 and τ_1 depends on the shape of the pulse required at the amplifier output.

Where it is only necessary for the amplitude to reach the maximum value $K_{\max}E$ during time T, the time constant τ_0 may be close to T. In this case the pulse shape is far from rectangular.

Where it is necessary to satisfactorily reproduce the pulse shape, the time constant τ_0 must be correlated with the leading edge output pulse time, while the time constant τ_1 must be large compared to the pulse duration T. This result is very important for the correct choice of parameters of a system transmitting discrete messages, since it defines the minimum time required for transition from one discrete level to another.

It should be noted that when amplifying a train of pulses, the above reasoning is valid, provided the pulse separation is wide enough to avoid the superposition of transient processes due to adjacent pulses.

Differentiating (6.13) with respect to t and equating E to unity, we obtain the following expression for the impulse response of an

aperiodic amplifier, which is shown in Fig. 6.4*a*:

$$g(t) = \frac{dv_1(t)}{dt} = -\frac{K_{max}}{\tau_0}\left(e^{-t/\tau_0} - \frac{\tau_0}{\tau_1}e^{-t/\tau_1}\right)$$

$$\approx -\frac{S}{C_0}\left(e^{-t/\tau_0} - \frac{\tau_0}{\tau_1}e^{-t/\tau_1}\right) \qquad (6.14)$$

Let us consider the transmission of a square pulse through an aperiodic transistor amplifier whose equivalent circuit is shown in

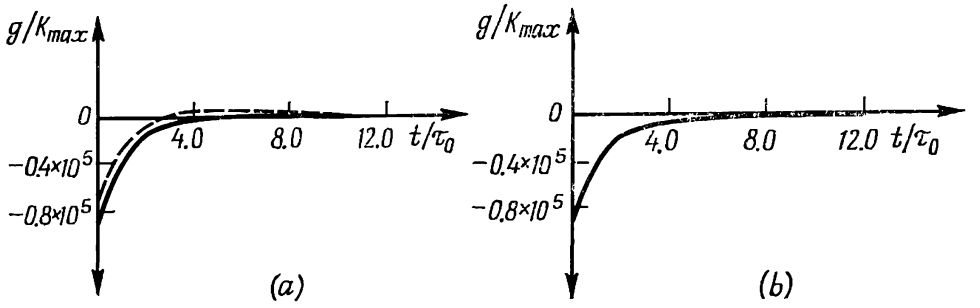


Fig. 6.4. Impulse response of a resistance-coupled amplifier
(*a*) with separation circuit R_1C_1 continuous line: $\tau_0 = 10^{-5}$ s, $\tau_1 = 10^{-2}$ s, $\tau_0/\tau_1 = 10^{-3}$; dashed line: $\tau_0 = 10^{-5}$ s, $\tau_1 = 10^{-4}$ s, $\tau_0/\tau_1 = 10^{-1}$; (*b*) without separation circuit

Fig. 5.14*a*. As it was noted in Sec. 5.6, for this purpose it is sufficient to let the capacitance C_1 go to infinity, i.e., to bridge the capacitor C_1, and include the conductance $G_1 = 1/R_1$ in G. In this case formula (6.13) transforms to

$$v_1(t) = -K_{max}E(1 - e^{-t/\tau_0}) \qquad (6.13')$$

(because $\tau_1 \to \infty$), and the impulse response becomes

$$g(t) = -(K_{max}/\tau_0)\,e^{-t/\tau_0} \qquad (6.14')$$

The pulse at the output of the amplifier in question is shown in Fig. 6.3*e*, and the impulse response, in Fig. 6.4*b*.

6.5. DIFFERENTIATION AND INTEGRATION OF SIGNALS

In radioelectronics it is often required to effect a signal conversion similar to differentiation or integration.

A signal $s(t)$ is applied to the input of a linear differentiating device so as to obtain an output signal

$$s_{out}(t) = \tau_0 \frac{ds(t)}{dt}$$

In an integrating circuit (or integrator), the relation between the output signal $s_{out}(t)$ and the input signal $s(t)$ is expressed by

$$s_{out}(t) = \frac{1}{\tau_0}\int s(t)\,dt$$

In these expressions τ_0 is a constant with the dimension of time.

Differentiation and integration are linear mathematical operations. Therefore, for differentiation or integration of signals one should use linear circuits and elements that have the required relationships between the input and output values. These requirements are in principle met by such elements as ordinary capacitors or inductors in combination with a resistor.

First, let us consider the circuit shown in Fig. 6.5*a*.

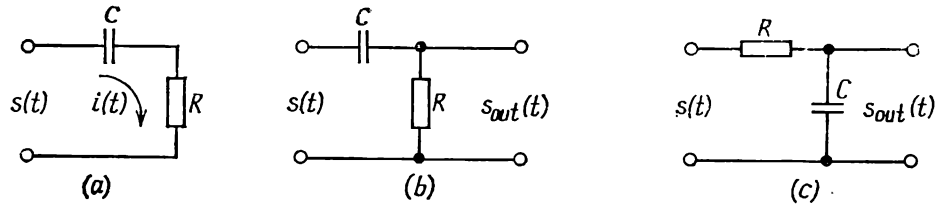


Fig. 6.5. (a) Circuit used for differentiation or integration, (*b*) differentiating *R-C* circuit and (*c*) integrating *R-C* circuit

Regarding the input signal $s(t)$ as an e.m.f., let us set up an equation for the circuit current $i(t)$:

$$Ri(t)+\frac{1}{C}\int i(t)\,dt=s(t) \tag{6.15}$$

Let this equation be multiplied by C and the time constant of the circuit be $\tau_0 = RC$.

Thus we have the equation

$$\tau_0 i(t)+\int i(t)\,dt=Cs(t) \tag{6.16}$$

The character of the functional relation between the current $i(t)$ and the input signal $s(t)$ depends on the value of the time constant τ_0. Let us consider two extreme cases of a very small and a very large τ_0.

In the first case, i.e., with very small τ_0, the first term of the left-hand side of equation (6.16) may be neglected. The differentiation of the remainder with respect to t yields

$$i(t)\approx C\frac{ds(t)}{dt}$$

From this it is evident that the voltage across the resistor R, coinciding in shape with $i(t)$, is proportional to the derivative of the input signal:

$$v_R=Ri(t)\approx RC\frac{ds(t)}{dt}=\tau_0\frac{ds(t)}{dt}$$

Thus, we arrive at the circuit diagram of a differentiating two-port network shown in Fig. 6.5*b* in which the output signal is taken from across the resistor R.

8163927

In the second case, i.e., with very large τ_0, the second term in the left-hand side of equation (6.16) may be omitted. In this case the current

$$i(t) \approx \frac{C}{\tau_0} s(t) = \frac{1}{R} s(t)$$

coincides in shape with the input signal, while the voltage across the capacitor C, equal to

$$v_C = \frac{1}{C} \int i(t)\, dt \approx \frac{1}{CR} \int s(t)\, dt$$

is proportional to the integral of the input signal $s(t)$. From this it follows that for integration purposes, the R-C circuit must be used

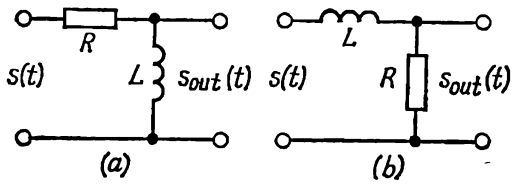


Fig. 6.6. (*a*) Differentiating R-L circuit and (*b*) integrating R-L circuit

in accordance with the arrangement shown in Fig. 6.5*c*. Similar results can be obtained by means of an R-L circuit (Fig. 6.6).

In a differentiating circuit, the time constant $\tau_0 = L/R$ must be sufficiently small, while in an integrating circuit, sufficiently large. The principle of differentiation with the first circuit (Fig. 6.6*a*) may be explained as follows. With an adequately high resistance R, the current through the R-L circuit is almost independent of L and coincides in shape with the input signal $s(t)$, while the output signal $s_{out}(t)$ taken from across the inductor L is given by

$$s_{out}(t) = L \frac{di}{dt} \approx L \frac{d}{dt} \left[\frac{1}{R} s(t) \right] = \tau_0 \frac{ds(t)}{dt}$$

Conversely, in the circuit shown in Fig. 6.6*b*, the current is defined mainly by the inductance L (since R is very low):

$$i(t) \approx \frac{1}{L} \int s(t)\, dt$$

while the output signal taken from across the resistor R is

$$s_{out}(t) = Ri(t) \approx \frac{1}{\tau_0} \int s(t)\, dt$$

Now let us give a more accurate definition of the above mentioned "small" and "large" τ_0. The simplest way is to analyse the spectrum. If the input signal $s(t)$ has a spectral density $S(\omega)$, then with exact differentiation, the output signal $s_{out}(t) = \tau_0\, ds(t)/dt$ must have a spectral density $i\omega\tau_0 S(\omega)$, while with exact integration, this signal

must have a spectral density $(1/i\omega\tau_0)\,\mathbf{S}(\omega)$ [see expressions (2.59) and (2.60)]. This means that for exact differentiation, it is necessary to use a two-port network with a transfer factor

$$\mathbf{K}(i\omega) = \tau_0\, i\omega \tag{6.17}$$

and for exact integration, a two-port network with a transfer factor

$$\mathbf{K}(i\omega) = 1/\tau_0 i\omega \tag{6.18}$$

Two two-port networks shown in Fig. 6.5*b*, and *c* have respectively transfer functions

$$\mathbf{K}(i\omega) = \frac{R}{R+1/i\omega C} = RC\frac{i\omega}{1+RCi\omega} = \frac{\tau_0 i\omega}{1+\tau_0 i\omega} \tag{6.19}$$

and

$$\mathbf{K}(i\omega) = \frac{1/i\omega C}{R+1/i\omega C} = \frac{1}{i\omega RC}\frac{1}{1+(1/i\omega RC)} = \frac{1}{\tau_0 i\omega}\frac{1}{1+(1/\tau_0 i\omega)} \tag{6.20}$$

Comparison between expressions (6.17) and (6.19) shows that for adequate differentiation, the following condition must be satisfied:

$$\tau_0\omega \ll 1 \tag{6.21}$$

This inequality must be satisfied for all frequencies of the input signal spectrum, including the uppermost frequency.

Comparison between expressions (6.18) and (6.20) shows that for adequate integration, the following condition must be satisfied:

$$\tau_0\omega \gg 1 \tag{6.22}$$

This inequality must be satisfied for all frequencies of the input signal spectrum, including the lowermost frequency.

From inequalities (6.21) and (6.22) it follows that the lower the frequency of the input signal, the higher the accuracy of differentiation of any specified circuit, while the accuracy of integration is higher for higher frequencies.

From these inequalities also stems the following principal proposition: the higher the accuracy of differentiation or integration, the lower the absolute value of the transfer function $\mathbf{K}(i\omega)$ of the circuit effecting this conversion of the signal. This applies to simple *R-C* or *R-L* circuits shown in Figs. 6.5 and 6.6. In the limit, with ideal conversion, $\mathbf{K}(i\omega) \to 0$.

Thus, simple *R-C* or *R-L* circuits are suitable only for approximate differentiation of signals. The accuracy of differentiation can in principle be increased by using an amplifier at the output of the differentiating circuit. However, the unavoidable instability of amplification and nonlinear distortion in a simple amplifier make such a method practically unacceptable. Therefore, precision differentiating devices use amplifiers with negative feedback, such as that shown in Fig. 6.7*a*.

The feedback voltage taken from across the resistor R_2 is connected into the circuit of the resistor R. Figure 6.7*b* shows an equivalent

circuit for the network located to the right of the terminals *2* and *2'*. The amplifier K_a having a high input impedance is considered here a dependent voltage-controlled voltage source. The feedback two-port network corresponds to the voltage divider R_1, R_2, so the transfer function of the feedback network is $K_{fb} = R_2/(R_1 + R_2)$.

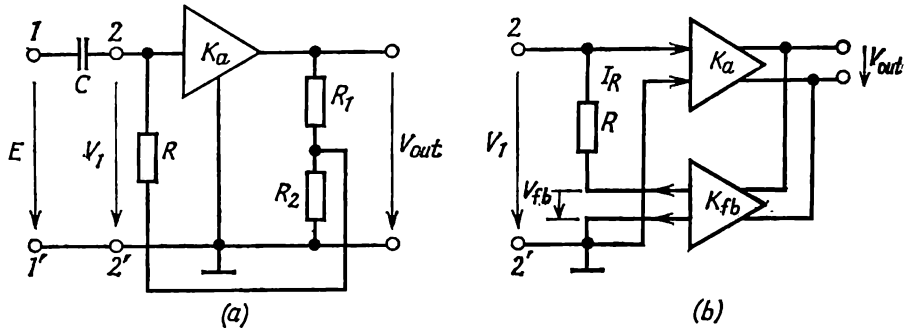


Fig.6.7. (*a*) Differentiating circuit with negative feedback and (*b*) equivalent circuit

The current through the resistor R is defined by the obvious expression $I_R = (V_1 - V_{fb})/R$. Taking into account that $V_{fb} = K_{fb}V_{out}$ and $V_{out} = K_aV_1$, we obtain

$$I_R = (V_1/R)(1 - K_aK_{fb}) = V_1/R_{eq}$$

where $R_{eq} = R(1 - K_aK_{fb})$ is the equivalent resistance across the terminals *2* and *2'*.

In the case of negative feedback, the condition $K_aK_{fb} < 0$ is satisfied. In the circuit under consideration, K_{fb} is a positive (real) number, while K_a is a negative number (for example, when using a common-emitter transistor amplifier, see Sec. 5.4).

Thus, the time constant of the circuit R_{eq}, C, defining the quality of differentiation, is

$$\tau_{eq} = R_{eq}C = RC/(1 + |K_aK_{fb}|) = \tau_0/(1 + |K_aK_{fb}|)$$

The transfer function of the device as a whole is given by

$$\mathrm{K}_0(i\omega) = \frac{\mathrm{V}_{out}}{\mathrm{E}} = \frac{\mathrm{V}_{out}}{\mathrm{V}_1}\frac{\mathrm{V}_1}{\mathrm{E}} = \frac{K_a}{1 + |K_aK_{fb}|}\frac{i\omega\tau_0}{1 + i\omega\tau_{eq}}$$

Since τ_{eq} can be made many times smaller than τ_0, the effect of the term $i\omega\tau_{eq}$ in the denominator of the second fraction can be reduced practically to zero (without reducing the magnitude of $\mathrm{K}_0(i\omega)$). For example, at $K_a = -100$ and $K_{fb} = 0.1$, $\tau_{eq} = \tau_0/(1 + 10)$ and $K_a\tau_{eq} = 100\tau_0/(1 + 10)$. As a result, we have

$$\mathrm{K}_0(i\omega) = 9\frac{i\omega\tau_0}{1 + i\omega\tau_0/11}$$

Modern precision differentiating devices employ *operational amplifiers* providing a very high gain to make it possible to effect any required approximation of the transfer function to the form $\mathbf{K}(i\omega) = ki\omega\tau_0$, where k is a constant.

In the above circuit (Fig. 6.7*a*), for $|K_aK_{fb}| \gg 1$ the transfer function

$$\mathbf{K}_0(i\omega) \approx \frac{1}{K_{fb}} \frac{i\omega\tau_0}{1+i\omega\tau_{eq}}$$

is almost independent of K_a. The ensuing advantages as to stable amplification and attenuation of nonlinear distortion were explained in Sec. 5.9.

In conclusion, let us find the impulse response of differenting and integrating circuits and give some examples of transmission of pulse signals through such circuits. It is easy to find the impulse response of an integrator.

Proceeding from the relation

$$s_{out}(t) = \frac{1}{\tau_0} \int_{-\infty}^{\infty} s(t)\, dt$$

and substituting the delta function $\delta(t)$ for $s(t)$, we obtain the following expression for $s_{out}(t)$, i.e., for the impulse response of an ideal integrator:

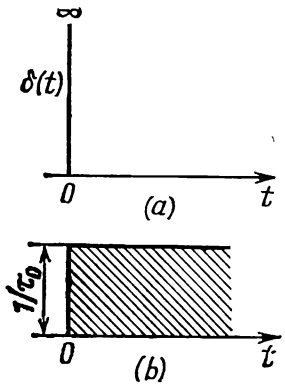


Fig. 6.8. (*a*) Unit impulse and (*b*) impulse response of an ideal integrating circuit

$$g(t) = \frac{1}{\tau_0} \int_{-\infty}^{\infty} \delta(t)\, dt = \frac{1}{\tau_0} \text{ for } 0 < t < \infty \tag{6.23}$$

The unit impulse and the impulse response of the integrator are shown in Fig. 6.8.

The impulse response of a simple integrating R-C circuit (Fig. 6.5*c*) is as follows

$$g(t) = \frac{1}{\tau_0} e^{-t/\tau_0} \tag{6.23'}$$

The determination of the impulse response of a differentiator is hindered by the necessity of finding the derivative of the delta function. This difficulty can be overcome if the short pulse which transforms to $\delta(t)$ as its duration τ approaches zero (see Sec. 2.11) is differentiated before proceeding to the limit. Figure 6.9*a* shows an initial triangular pulse with a base of 2τ and a height of $1/\tau$. The pulse area is equal to unity. The derivative of such a function is illustrated in Fig. 6.9*b*. As $\tau \to 0$, the triangular pulse transforms to the delta function $\delta(t)$, while the double bipolar pulse (Fir. 6.9*b*) transforms to the derivative of the delta function, i.e., $\delta'(t)$.

Thus, the impulse response of an ideal differentiator, obtained from the general expression $s_{out}(t) = \tau_0 s'(t)$ by replacing $s(t)$ by $\delta(t)$ and $s_{out}(t)$ by $g(t)$, is defined by the expression $g(t) = \tau_0 \delta'(t)$ for $0 < t < \infty$. It is shown in Fig. 6.9*b* for $\tau \to 0$.

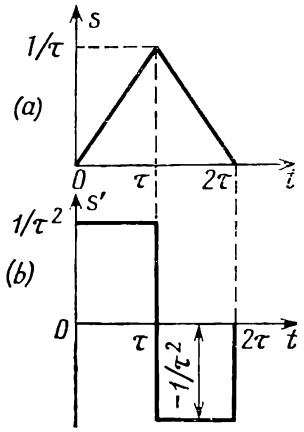


Fig. 6.9. Determining the impulse response of an ideal differentiating circuit

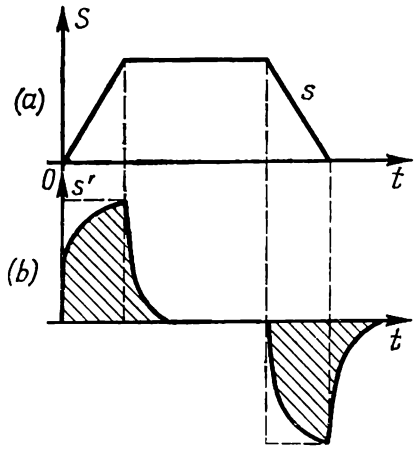


Fig. 6.10. Signal at (*a*) the input and (*b*) the output of a differentiating circuit

Figure 6.10 illustrates the transmission of a trapezoidal pulse through a differentiating *R-C* circuit. The dashed lines show the signal at the output of an ideal differentiator.

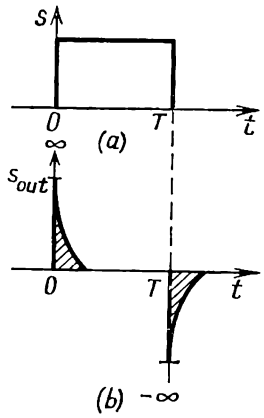


Fig. 6.11. Signal at (*a*) the input and (*b*) the output of a differentiating circuit

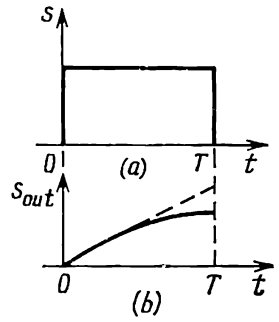


Fig. 6.12. Signal at (*a*) the input and (*b*) the output of an integrating circuit

In Fig. 6.11, similar graphs are drawn for a square-pulse input signal (Fig. 6.11*a*). With exact differentiation, the output signal must have the shape of two unit impulses: $\delta(t)$ and $-\delta(t - T)$.

In reality, we have two exponential pulses (shaded in Fig. 6.11*b*).

The examples given in Figs. 6.10 and 6.11 indicate that the slower the input signal changes with time, the better the differentiation.

Figure 6.12 shows the operation of an integrating *R-C* circuit whose input is fed with a square pulse. The larger the time constant of the circuit, the closer the actual output signal (continuous line) to the ideal one (dashed line).

6.6. ANALYSIS OF RADIO SIGNALS IN SELECTIVE CIRCUITS. APPROXIMATE SPECTRAL METHOD

The problems discussed in the preceding chapter have one feature in common—they deal with signals whose shape coincides with the message being transmitted. When transmitting such signals, the problem of preserving the information is closely associated with that of preserving the shape of the signals.

It is different with a radio signal in which the information is embedded in one of the several parameters of a high-frequency oscillation. The structure of this oscillation need not be completely preserved; it is sufficient to keep intact the law govering the variation of the information-containing parameter. Thus, in the case of an amplitude-modulated oscillation, it is necessary to transmit the envelope, whereas a slight change in the frequency or phase of the carrier, having no practical importance, can be disregarded in the analysis. On the other hand, in the case of angle-modulated radio signals, our attention must be focused on the exact reproduction of the law governing frequency and phase variation.

These specific features of radio signals allow one to simplify to a certain degree the analysis of their transmission through linear systems. This simplification is especially useful when a radio signal is a narrow-band process, and the circuit is a narrow-band system. These conditions are typical of actual radio signals and selective circuits. In Sec 3.1 we have already mentioned that even in the case of "wide-band" signals, the spectrum of a radio signal is small compared to its carrier frequency. Correspondingly, the passband of the circuit is usually narrow in comparison with its resonance frequency.

The spectral density $S_a(\omega)$ of a modulated high-frequency wave $a(t)$ forms two spikes near the frequencies ω_0 and $-\omega_0$, while the transfer function $\mathbf{K}(i\omega)$ forms two spikes near the frequencies ω_r and $-\omega_r$ (Fig. 6.13). For generalization purposes, it is assumed here that the resonance frequency ω_r of the circuit does not coincide with the centre frequency ω_0 of the signal, the detuning

$$\Delta\omega = \omega_0 - \omega_r \tag{6.24}$$

being of the same order of magnitude as the passband of the circuit.

Let us set up an expression for the output signal of the circuit. If the input signal can be represented in the form $a(t) = A(t)\cos[\omega_0 t + \theta(t)]$, calculations will be much simplified when using the analytic signal [see Sec. 3.10, formulas (3.87) and (3.88)]

$$z(t) = \mathbf{A}(t)\, e^{i\omega_0 t} \tag{6.25}$$

The spectral density [modulus $Z(\omega)$] of this signal is shown in Fig. 6.13 by a heavy line (compare with Fig. 3.25). Since the function

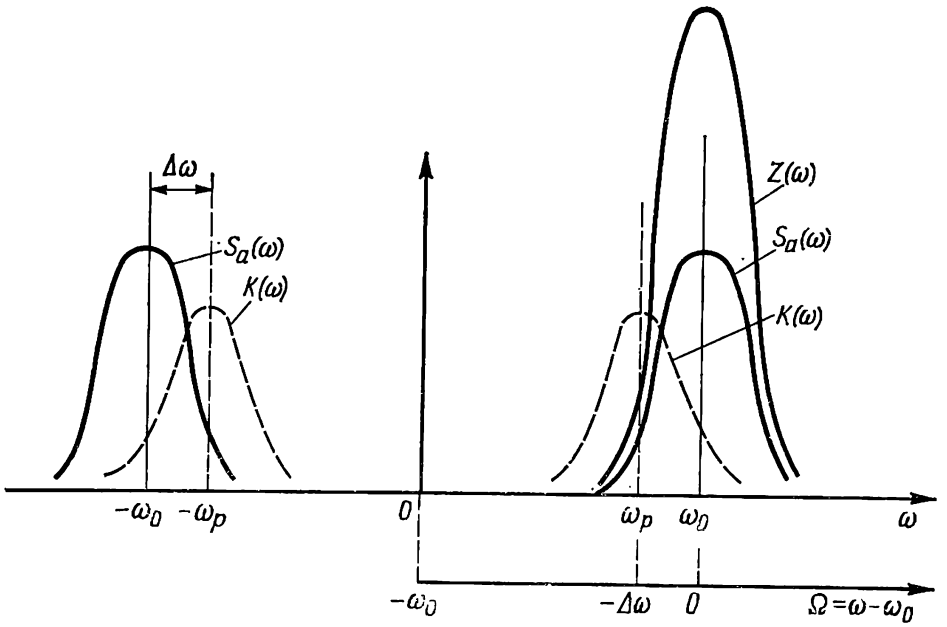


Fig. 6.13. Spectral density of a modulated oscillation and the transfer function of a narrow-band circuit

$\mathbf{Z}(\omega)$ exists only in the positive-frequency region, the analytic signal at the output of the system should be found from the expression

$$z_{out}(t) = \frac{1}{2\pi}\int_0^\infty \mathbf{Z}(\omega)\,\mathbf{K}(i\omega)\, e^{i\omega t}\, d\omega \tag{6.26}$$

In Sec. 3.10 it was shown that $\mathbf{Z}(\omega) = 2\mathbf{S}_a(\omega)$ for $\omega > 0$, $\mathbf{S}_a(\omega)$ being equal to $\frac{1}{2}\mathbf{S}_A(\omega - \omega_0)$ in the positive-frequency region [see formula (3.10) derived for the special case $\theta(t) = \theta_0$; when using complex envelope, the latter includes $\theta(t)$].

Hence, $\mathbf{Z}(\omega) = \mathbf{S}_A(\omega - \omega_0)$. Substituting this expression into (6.26), we obtain

$$z_{out}(t) = \frac{1}{2\pi} \int_0^\infty \mathbf{S}_A(\omega - \omega_0)\, \mathbf{K}(i\omega)\, e^{i\omega t}\, d\omega \qquad (6.27)$$

Now, let us introduce a new variable $\Omega = \omega - \omega_0$. Then

$$z_{out}(t) = \left\{ \frac{1}{2\pi} \int_{-\omega_0}^\infty \mathbf{S}_A(\Omega)\, \mathbf{K}[i(\omega_0 + \Omega)]\, e^{i\Omega t}\, d\Omega \right\} e^{i\omega_0 t} \qquad (6.28)$$

From comparison of this expression with (6.25), it becomes immediately obvious that the expression in braces corresponds to the complex envelope of the output oscillation

$$\mathbf{A}_{out}(t) = A(t)_{out}\, e^{i\theta_{out}(t)} = \frac{1}{2\pi} \int_{-\omega_0}^\infty \mathbf{S}_A(\Omega)\, \mathbf{K}[i(\omega_0 + \Omega)]\, e^{i\Omega t}\, d\Omega \qquad (6.29)$$

Further simplification of the analysis stems from the properties of the transfer function of resonance circuits that feature a pronounced frequency selectivity. Here, the magnitude of the transfer factor $\mathbf{K}(i\omega)$ drops rapidly as ω deviates farther away from the resonance frequency.

Therefore, it is advisable to express the transfer function as a function of deviation of the frequency ω from the resonance frequency ω_r:

$$\mathbf{K}(i\omega) = \mathbf{K}[i(\omega_r + \nu)] = \mathbf{K}_1(i\omega) = \mathbf{K}_1[i(\Delta\omega + \Omega)] \qquad (6.30)$$

where $\nu = \omega - \omega_r = \Delta\omega + \Omega$, and $\Delta\omega = \omega_0 - \omega_r$ is the constant detuning parameter.

Since for $\Omega = -\omega_0$ the transfer factor $\mathbf{K}_1 = [i(\Delta\omega + \Omega)]$ is practically zero, the lower limit of the integral in expression (6.29) may be replaced by $-\infty$. Then, expression (6.29) assumes the following form:

$$\mathbf{A}_{out}(t) = \frac{1}{2\pi} \int_{-\infty}^\infty \mathbf{S}_A(\Omega)\, \mathbf{K}_1[i(\Delta\omega + \Omega)]\, e^{i\Omega t}\, d\Omega \qquad (6.31)$$

This expression in no way differs from the conventional Fourier integral defining the original signal in terms of the given spectral density $\mathbf{S}_A(\Omega)$ of the signal envelope and the transfer function $\mathbf{K}_1[i(\Delta\omega + \Omega)]$.

Replacing $i\Omega$ by p, we get the above expression in the form of an inverse Laplace transform:

$$\mathbf{A}_{out}(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \overline{\mathbf{S}}_A(p)\, \mathbf{K}_1[i\Delta\omega + p]\, e^{pt}\, dp \qquad (6.32)$$

Thus, the analysis transmission of a narrow-band high-frequency signal through a selective circuit is reduced essentially to the analy-

sis of the changes suffered by the complex envelope of the signal. After finding $A_{out}(t)$ and $\theta_{out}(t)$ for the output (analytic) signal, we can write the following expression:

$$z_{out}(t) = A_{out}(t)\, e^{i[\omega_0 t + \theta_{out}(t)]} \tag{6.33}$$

whence

$$a_{out}(t) = A_{out}(t) \cos[\omega_0 t + \theta_{out}(t)] \tag{6.34}$$

The calculations involved in finding $\mathbf{A}_{out}(t)$ by formula (6.32) are simpler than those required for direct computation of $a_{out}(t)$ by the inverse Laplace transformation, because the transformation between $\mathbf{S}_a(\omega)$ and $\mathbf{S}_A(\Omega)$, and between $\mathbf{K}(p)$ and $\mathbf{K}_1(i\Delta\omega + p)$, decreases the number of singular points of the integrand.

6.7. SIMPLIFICATION OF THE SUPERPOSITION INTEGRAL METHOD (ENVELOPE METHOD)

In the preceding section, simplification of the spectral method was attained by simplifying the transfer function $\mathbf{K}(i\omega)$ of the selective circuit. Similarly, the superposition integral method can be simplified by simplifying the impulse response $g(t)$ that is closely related to the transfer function $\mathbf{K}(i\omega)$.

Proceeding from the general expression (5.28)

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{K}(i\omega)\, e^{i\omega t}\, d\omega$$

and transforming to the analytic signal $z_g(t)$ that corresponds to the physical function $g(t)$, we may write [see (3.91)]

$$z_g(t) = \frac{1}{\pi} \int_{0}^{\infty} \mathbf{K}(i\omega)\, e^{i\omega t}\, d\omega \tag{6.35}$$

As in the preceding section, let us replace the integration variable $\omega = \omega_r + \nu$. Then, in view of formula (6.30), we get

$$z_g(t) = \left[\frac{1}{\pi} \int_{-\infty}^{\infty} \mathbf{K}_1(i\nu)\, e^{i\nu t}\, d\nu\right] e^{i\omega_r t} \tag{6.36}$$

On the other hand, representing the sought impulse response in the form

$$g(t) = G(t) \cos[\omega_r t + \gamma_g(t)]$$

we have

$$z_g(t) = G(t)\, e^{i[\omega_r t + \gamma_g(t)]} = G(t)\, e^{i\gamma_g(t)} e^{i\omega_r t} = \mathbf{G}(t)\, e^{i\omega_r t} \tag{6.37}$$

Comparison between expressions (6.36) and (6.37) directly yields an equation determining *the complex envelope of the impulse response*

$g(t)$:

$$\mathbf{G}(t) = G(t)\, e^{i\gamma_g(t)} = 2\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{K}_1(i\nu)]\, e^{i\nu t} d\nu \qquad (6.38)$$

As we shall see from examples given below, the use of this expression simplifies the calculation of the impulse response $g(t)$.

Referring now to expression (6.31) and using formula (2.64), we may define $\mathbf{A}_{out}(t)$ as the convolution of two time functions corresponding to the spectral functions $\mathbf{S}_A(\Omega)$ and $\mathbf{K}_1(i\nu)$.

The first of these spectral functions corresponds to $\mathbf{A}(t)$ and the second, to the function $\mathbf{G}(t)/2$, as it follows from (6.38).

Therefore

$$\mathbf{A}_{out}(t) = \frac{1}{2} \int_{-\infty}^{\infty} \mathbf{A}(x)\, \mathbf{G}(t-x)\, dx$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} A(x)\, G(t-x)\, e^{i[\theta(x)+\gamma_g(t-x)]}\, dx \qquad (6.39)$$

This is a general expression suitable for any selective circuits and any narrow-band signals. In cases where free oscillations are characterized by a constant carrier frequency (for example, in single-tuned circuits), $\gamma_g(t)$ degenerates into a constant phase, and expression (6.39) is considerably simplified. The same applies to the case of an unmodulated carrier, where $\theta(t)$ becomes a constant.

The superposition integral method is more effective where the time characteristics of signals or circuits (or both) are simpler than the spectral characteristics, as in particular, is the case with some frequency-modulated signals. The application of the envelope method is exemplified in Sec. 6.10.

6.8. TRANSMISSION OF A RADIO PULSE THROUGH A RESONANCE AMPLIFIER

Having in mind a radio pulse with a rectangular envelope and unmodulated high-frequency carrier, let us first consider the phenomena occurring in the resonance amplifier circuit shown in Fig. 5.17 when transmitting the leading edge of the pulse, i.e., when switching on a harmonic e.m.f. $e(t) = E_0 \cos(\omega_0 t + \theta_0)$ at an instant $t = 0$. The voltage across the tank (i.e., resonant) circuit of the amplifier is considered to be the output quantity.

First, let us derive an exact expression for the output voltage.

Using the general formula (5.60), let us multiply the numerator and denominator of the fraction in this expression by $i\omega/C$:

$$\mathbf{K}(i\omega) = -Si\omega/C \left[\frac{i\omega(G_l + G_{sh})}{C} + (i\omega)^2 + \frac{1}{LC}\right]$$

Let us use the well-known relations: $1/\sqrt{LC} = \omega_r$—the resonance frequency of the tank circuit: $(G_i + G_{sh})/2C = \alpha_t = 1/\tau_t$, where τ_t is the time constant of the tank circuit and α_t is the attenuation.

Furthermore, let us use expression (5.61) for the resonance ratio $K_{\max}$. Then the transfer function of the amplifier can be reduced to the form

$$\mathbf{K}(i\omega) = -2\alpha_t K_{\max} \frac{i\omega}{(i\omega)^2 + 2\alpha_t i\omega + \omega_r^2} \tag{6.40}$$

Thus, in operator notation we have

$$\mathbf{K}(p) = -2\alpha_t K_{\max} \frac{p}{p^2 + 2\alpha_t p + \omega_r^2} \tag{6.41}$$

The Laplace transform of the oscillation $E_0 \cos(\omega_0 t + \theta_0)$, $t \geqslant 0$, has the following form:

$$\overline{\mathbf{E}}(p) = \frac{1}{2} e^{i\theta_0} \frac{E_0}{p - i\omega_0} + \frac{1}{2} e^{-i\theta_0} \frac{E_0}{p + i\omega_0} \tag{6.42}$$

The voltage at the amplifier output [see formula (6.3)] is

$$v(t) = -2\alpha_t K_{\max} E_0 \left\{ \frac{1}{2} e^{i\theta_0} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{pe^{pt}\,dp}{(p - i\omega_0)(p^2 + 2\alpha_t p + \omega_r^2)} \right.$$

$$\left. + \frac{1}{2} e^{-i\theta_0} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{pe^{pt}\,dp}{(p + i\omega_0)(p^2 + 2\alpha_t p + \omega_r^2)} \right\} \tag{6.43}$$

Calculating the residues at the four poles, we obtain the following final expression:

$$v(t) = -V \cos(\omega_0 t + \theta_0 - \varphi) + Ve^{-t/\tau_t} \left\{ \cos(\theta_0 - \varphi) \cos \omega_{fr}(t) \right.$$

$$\left. - \left[\frac{\omega_r^2}{\omega_0 \omega_{fr}} \sin(\theta_0 - \varphi) + \frac{\alpha_t}{\omega_{fr}} \cos(\theta_0 - \varphi) \right] \sin \omega_{fr} t \right\} \tag{6.44}$$

where $\omega_{fr} = \sqrt{\omega_r^2 - \alpha_t^2}$ is the frequency of free oscillations and $\varphi = \arctan(\omega_0 - \omega_r)\,\tau_t$ is the phase shift (under steady-state conditions). The first addend in (6.44) defines the steady-state oscillation and the second, the free oscillation.

Now let us use approximate expression (6.32). The transfer function is found from formula (5.64), in which, in this case, by a_{eq} is meant

$$a_{eq} = \frac{2(\omega - \omega_r)}{\omega_r} Q_{eq} = \frac{2[(\omega_0 + \Omega) - \omega_r]}{\omega_r} Q_{eq} = (\Delta\omega + \Omega)\,\tau_t$$

where $Q_{eq} = \omega_r/2\alpha_t = \omega_r \tau_t/2$ is the Q-factor of the circuit with account being taken of the conductivity G_i of the active element.

Thus,

$$\mathbf{K}_1[i(\Delta\omega+\Omega)] = -\frac{K_{\max}}{1+i(\Delta\omega+\Omega)\tau_t} \tag{6.45}$$

$$\mathbf{K}_1(i\Delta\omega+p) = -\frac{K_{\max}}{1+(i\Delta\omega+p)\tau_t} \tag{6.46}$$

Let us now set up a simplified expression for $\bar{\mathbf{S}}_A(p)$. In the given case (with an unmodulated high-frequency carrier), the envelope $\mathbf{A}(t)$ is a real function and has the form of a step E_0. The spectral density of this envelope is given by

$$\mathbf{S}_A(\Omega) = E_0/i\Omega$$

Taking into account the initial phase θ_0 of the oscillations, the spectral density of the complex envelope $\mathbf{A}(t)$ may be written in the form

$$\mathbf{S}_A(\Omega) = e^{i\theta_0}E_0/i\Omega$$

and its Laplace transform

$$\bar{\mathbf{S}}_A(p) = e^{i\theta_0}E_0/p \tag{6.47}$$

Substituting formulas (6.46) and (6.47) into (6.32), we find the complex envelope at the amplifier output:

$$\mathbf{A}_{out}(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\bar{\mathbf{S}}_A(p)\,\mathbf{K}_1[i\Delta\omega+p]\,e^{pt}\,dp$$

$$= -K_{\max}E_0e^{i\theta_0}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{e^{pt}\,dp}{p[1+(i\Delta\omega+p)\tau_t]} \tag{6.48}$$

The integrand has only two poles:

$$p_1 = 0, \quad p_2 = -(1+i\Delta\omega\tau_t)/\tau_t \tag{6.49}$$

The residues at these poles

$$\mathrm{res}_1 = \left[\frac{e^{pt}}{1+i\Delta\omega\tau_t+2\tau_t p}\right]_{p=p_1} = \frac{1}{1+i\Delta\omega\tau_t} \tag{6.50}$$

$$\mathrm{res}_2 = \left[\frac{e^{pt}}{1+i\Delta\omega\tau_t+2\tau_t p}\right]_{p=p_2} = -\frac{e^{-(1/\tau_t+i\Delta\omega)t}}{1+i\Delta\omega\tau_t} \tag{6.51}$$

The sum of the residues is

$$\mathrm{res}_1+\mathrm{res}_2 = \frac{1-e^{-(1/\tau_t+i\Delta\omega)t}}{1+i\Delta\omega\tau_t} = \frac{1-e^{-t/\tau_t}(\cos\Delta\omega t - i\sin\Delta\omega t)}{\sqrt{1+\Delta\omega^2\tau_t^2}\,e^{i\varphi}}$$

$$= \frac{\sqrt{1-2e^{-t/\tau_t}\cos\Delta\omega t+e^{-2t/\tau_t}}}{\sqrt{1+\Delta\omega^2\tau_t^2}}\,e^{i[\xi(t)-\varphi]} \tag{6.52}$$

where

$$\varphi = \arctan\Delta\omega\tau_t \tag{6.53}$$

is the phase shift of the voltage across the tank circuit with respect to the input e.m.f. under steady-state conditions, and

$$\xi(t) = \arctan \frac{e^{-t/\tau_t} \sin \Delta\omega t}{1 - e^{-t/\tau_t} \cos \Delta\omega t} \tag{6.54}$$

Thus, the complex envelope of the output voltage is

$$\mathbf{A}_{out}(t) = -\frac{K_{max} E_0}{\sqrt{1+\Delta\omega^2\tau_t^2}} \sqrt{1 - 2e^{-t/\tau_t} \cos \Delta\omega t + e^{-?t/\tau_t}}\, e^{i[\xi(t)+\theta_0-\varphi]} \tag{6.55}$$

and, after isolating the real part, the instantaneous voltage

$$v(t) = -\frac{K_{max} E_0}{\sqrt{1+\Delta\omega^2\tau_t^2}} \times \sqrt{1 - 2e^{-t/\tau_t} \cos \Delta\omega t + e^{-2t/\tau_t}} \cos [\omega_0 t + \theta_0 - \varphi + \xi(t)] \tag{6.56}$$

To compare the obtained result with the exact expression (6.44), let us reduce $v(t)$ to the form of a sum of two oscillations—forced and free. To this end, let us return to formulas (6.48) through (6.55) and set up an expression for the analytic signal corresponding to the voltage $v(t)$:

$$z_v(t) = \mathbf{A}_{out}(t)\, e^{i\omega_0 t} = -K_{max} E_0 \frac{1 - e^{-(1/\tau_t + \Delta\omega)t}}{1 + i\Delta\omega\tau_t} e^{i(\omega_0 t + \theta_0)}$$

$$= -K_{max} E_0 \frac{e^{i(\omega_0 t+\theta_0)} - e^{-t/\tau_t} e^{i[(\omega_0-\Delta\omega)t+\theta_0]}}{1 + i\Delta\omega\tau_t} \tag{6.57}$$

Reducing this expression to a trigonometric form, with due regard for (6.53) and the equality $\omega_0 - \Delta\omega = \omega_r$, we obtain

$$v(t) = -\frac{K_{max} E_0}{\sqrt{1+\Delta\omega^2\tau_t^2}} [\cos(\omega_0 t + \theta_0 - \varphi) - e^{-t/\tau_t} \cos(\omega_r t + \theta_0 - \varphi)]$$

192 (6.58)

For $\alpha_t/\omega_r \ll 1$, $\omega_{fr} \to \omega_r$ and expressions (6.44) and (6.58) are practically the same.

Let us consider the important practical results stemming from expression (6.56). First, let us discuss the case where the tank circuit is accurately tuned to the frequency of the exciting e.m.f. Equating ω_r to ω_0, we obtain $\Delta\omega = 0$. In this case, expression (6.56) is simplified:

$$v(t) = -K_{max} E_0 (1 - e^{-t/\tau_t}) \cos(\omega_0 t + \theta_0) \tag{6.59}$$

From this expression it follows that when the frequencies ω_0 and ω_r coincide, the envelope of the output oscillation rises according to $1 - e^{-t/\tau_t}$ irrespective of the phase of the e.m.f. at the moment it is switched on.

The corresponding curve calculated by the formula

$$A_{out}(t)/K_{\max}E_0 = 1 - e^{-t/\tau_t} \tag{6.60}$$

is plotted in Fig. 6.14. The same figure shows the graphs of the function

$$\frac{A_{out}(t)}{K_{\max}E_0} = \frac{1}{\sqrt{1+\Delta\omega^2\tau_t^2}}\sqrt{1-2e^{-t/\tau_t}\cos\Delta\omega t + e^{-2t/\tau_t}} \tag{6.61}$$

calculated for two values of the detuning parameter $\Delta\omega\tau_t$, equal to 1 and 2.

From Fig. 6.14 it is clear that when the detuning is large the response reaches its steady-state in an oscillatory manner. This is due to the beating together of two oscillations of frequencies ω_0 and ω_{fr}, the latter, under the above assumption of the high Q of the tank circuit, differing but very slightly from the resonance frequency ω_r.

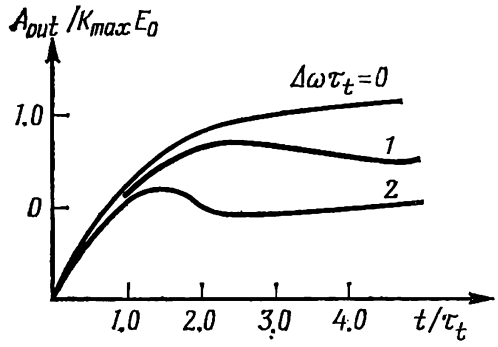


Fig. 6.14. Process of setting in of the envelope of an r-f voltage at the output of a resonance amplifier acted upon by a harmonic e.m.f. for different detuning parameters

Shown in Fig. 6.15 are the graphs of the normalized envelope, i.e., the graphs of the function $A_{out}(t)\sqrt{1+\Delta\omega^2\tau_t^2}:K_{\max}E_0$. It is clear that as the detuning increases, the steepness of the leading edge of the envelope also increases and the transition time is somewhat reduced. The graphs of the function $\xi(t)$, i.e., of the variable part $\theta_{out}(t)$, are given in Fig. 6.16.

Let us use the obtained results to determine the shape and parameters of a radio pulse at the output of a single-tuned amplifier excited by a pulse of rectangular envelope.

The oscillation at the input (Fig. 6.17*a*) is defined by the expression

$$a(t) = \begin{cases} E_0\cos(\omega_0 t + \theta_0) & \text{for } 0 < t < T \\ 0 & \text{for } t < 0 \text{ and } t > T \end{cases} \tag{6.62}$$

As in Sec. 6.4, the problem can be solved by studying independently the phenomena occurring during the rise and decay of the output pulse and then superposing the solutions obtained.

If the pulse duration T is longer than the actual transition time for the tank circuit in the case of harmonic excitation, then by the instant of termination of the input pulse, the amplitude of the oscil-

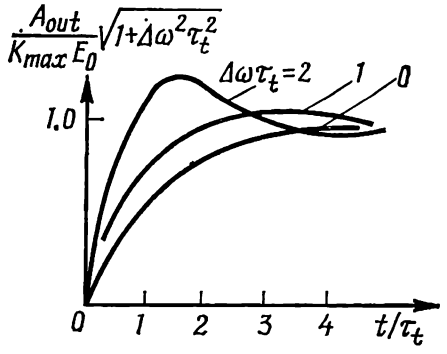


Fig. 6.15. The same as in Fig. 6.14, but with the envelope normalized with respect to the steady-state value

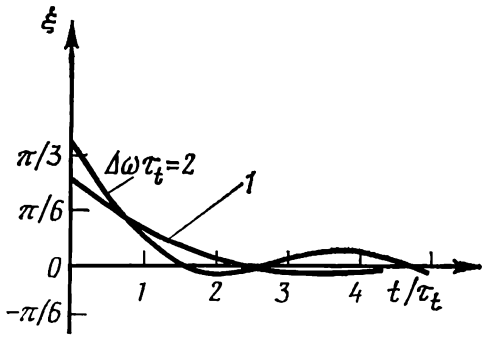


Fig. 6.16. Process of setting in of the phase of an oscillation as a function of the detuning parameter $\Delta\omega\tau_t$

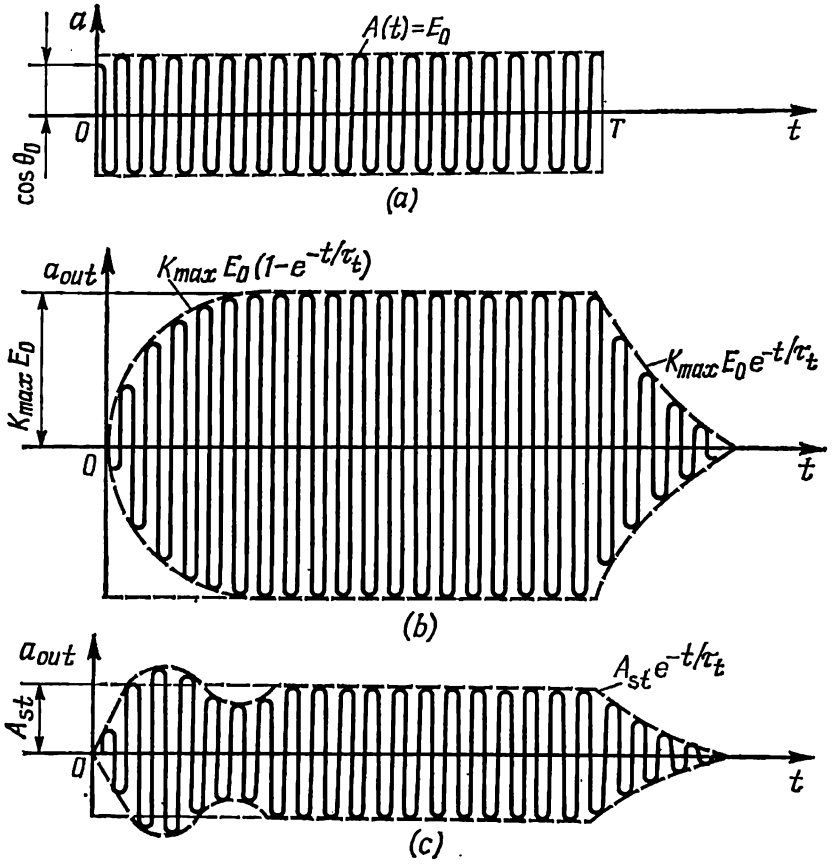


Fig. 6.17. Transmission of a radio pulse through a resonance amplifier
(*a*) the pulse at the amplifier input; (*b*) the output pulse with the tank circuit being accurately tuned; (*c*) the output pulse in the case of detuning

lations at the amplifier output will reach its steady-state value.

$$A_{out,\,st} = K_{\max} E_0 / \sqrt{1 + \Delta\omega^2 \tau_i^2} = \text{const}$$

Starting from the time $t = T$, i.e., after the exciting e.m.f. has ceased, only a free oscillation remains, which can be represented in the form

$$a_{out}(t) = A_{out,\,st} e^{-t/\tau_t} \cos(\omega_{fr} t + \varphi_0) \approx \frac{K_{\max} E_0}{\sqrt{1 + \Delta\omega^2 \tau_i^2}} e^{-t/\tau_t} \cos(\omega_r t + \varphi_0) \quad \text{for } t > T \qquad (6.63)$$

Here, φ_0 stands for the phase of the voltage across the tank circuit at the instant $t = T$; $\omega_r \approx \omega_{fr}$.

Thus, in contrast to the leading edge of the pulse, the trailing edge has the shape of an exponential curve regardless of the relation between the frequencies ω_0 and ω_r. The signal at the output of the amplifier with $\Delta\omega\tau_t = 0$ and $\Delta\omega\tau_t = 2$ (Fig. 6.17*b* and *c*) is shown for the case where the pulse duration is much longer than the transition time.

6.9. LINEAR DISTORTION OF OSCILLATIONS WITH CONTINUOUS AMPLITUDE MODULATION

Applied at an instant $t = 0$ to the input of the single-tuned amplifier shown in Fig. 5.17 is an e.m.f.

$$a(t) = E_0 [1 + M \cos(\Omega t + \gamma_0)] \cos(\omega_0 t + \theta_0) \qquad (6.64)$$

The epoch angle of the modulating function is $\gamma = \gamma_0 = \text{const}$.

Let us find the structure of the voltage at the amplifier output under both transient and steady-state conditions.

In this case, it is expedient to use the superposition integral method [see expression (6.39)]. In accordance with (6.64), the complex envelope of the oscillation is

$$\mathbf{A}(t) = E_0 [(1 + M \cos(\Omega t + \gamma_0)] e^{i\theta_0} \qquad (6.65)$$

Let us find the complex envelope of the impulse response of the circuit shown in Fig. 5.17. According to formula (6.38) in operator notation, the complex envelope is given by

$$\mathbf{G}(t) = 2 \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \mathbf{K}_1(p)\, e^{pt}\, dp \qquad (6.66)$$

Substituting into this expression the transfer function by formulas (6.45)—(6.46) and considering that $\Delta\omega + \Omega = \nu$, we get

$$\mathbf{G}(t) = -2K_{\max} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{pt}\, dp}{\sqrt{1 + p\tau_t}}$$

From formula (6.51) it follows that the residue at the pole $p = -1/\tau_t$ is

$$\text{res} = e^{-t/\tau_t}$$

and therefore,

$$\mathbf{G}(t) = -(2K_{\max}/\tau_t)\, e^{-t/\tau_t} \tag{6.67}$$

It is worthy of notice that the analytic signal corresponding to the impulse response $g(t)$ has the form

$$z_g(t) = \mathbf{G}(t)\, e^{i\omega_r t} = -\frac{2K_{\max}}{\tau_t}\, e^{-t/\tau_t} e^{i\omega_r t} \tag{6.68}$$

while the impulse response itself is

$$g(t) = \text{Re} z_g(t) = -\frac{2K_{\max}}{\tau_t}\, e^{-t/\tau_t} \cos \omega_r t \tag{6.69}$$

Now, let us analyze the most interesting case of exact tuning of the amplifier to the carrier frequency of the modulated wave ($\omega_0 = \omega_r$, $\Delta\omega = 0$). Referring now to expression (6.39) and substituting into it for mulas (6.65) and (6.67), and also, taking into account that in the given case $\theta(t) = \theta_0 = \text{const}$, $\gamma_g(t) = 0$, we have

$$\mathbf{A}_{out}(t) = -\frac{1}{2} E_0 \frac{2K_{\max}}{\tau_t} e^{i\theta_0} \int_0^t [1 + M \cos(\Omega x + \gamma_0)]\, e^{-(t-x)/\tau_t}\, dx \tag{6.70}$$

The limits of integration are matched to the beginning ($t = 0$) of the action of the external force and to the property of the impulse response to vanish for $(t - x) < 0$.

The calculation of the simple integral in (6.70) results in the following expression for the complex envelope of the output voltage:

$$\mathbf{A}_{out}(t) = -K_{\max} E_0 \left\{1 + \frac{M}{\sqrt{1+\Omega^2\tau_t^2}} \cos(\Omega t + \gamma_0 - \xi_0) - e^{-t/\tau_t}\left[1 + \frac{M}{\sqrt{1+\Omega^2\tau_t^2}} \cos(\gamma_0 - \xi_0)\right]\right\} e^{i\theta_0} \tag{6.71}$$

where

$$\xi_0 = \text{arc tan}\, \Omega\tau_t = \text{arc tan}\, a_{eq} \tag{6.72}$$

Thus, the instantaneous value of the output voltage is

$$a_{out}(t) = -K_{\max} E_0 \left\{1 + \frac{M}{\sqrt{1+\Omega^2\tau_t^2}} \cos(\Omega t + \gamma_0 - \xi_0) - e^{-t/\tau_t}\left[1 + \frac{M}{\sqrt{1+\Omega^2\tau_t^2}} \cos(\gamma_0 - \xi_0)\right]\right\} \cos(\omega_0 t + \theta_0) \tag{6.73}$$

Let us compare this expression with (6.64). As would be expected, the frequency and phase of the amplitude-modulated wave passing through a resonance amplifier (with $\omega_0 = \omega_r$) are not changed.

The time-lag nature of the oscillatory circuit influences the *rate of change of the envelope in time*. This factor manifests itself under both transient and steady-state conditions.

Under transient conditions, the time-lag nature of the circuit results in that the envelope of the output signal always starts from zero, irrespective of the envelope value of the input e.m.f. at the moment it is switched on, i.e., irrespective of the epoch angle γ_0. The behaviour of this circuit relative to the envelope is the same as that of an aperiodic circuit with a time constant τ_t relative to a low-frequency voltage of frequency Ω.

Under steady-state conditions (when $t \gg \tau_t$), the output oscillation has the following form:

$$a_{out,\,st}(t) = -K_{\max}E_0\left[1+\frac{M}{\sqrt{1+\Omega^2\tau_t^2}}\cos(\Omega t+\gamma_0-\xi_0)\right]\cos(\omega_0 t+\theta_0) \tag{6.74}$$

The envelope of this oscillation differs from the envelope of the input oscillation in two respects:

1. The depth of modulation at the output

$$M_{out} = M/\sqrt{1+\Omega^2\tau_t^2} = M/\sqrt{1+a_{eq}^2}$$

is shallower than at the input, the relative decrease in the depth of modulation being

$$D = \frac{M_{out}}{M} = \frac{1}{\sqrt{1+a_{eq}^2}} = \frac{1}{\sqrt{1+(2\Omega Q_{eq}/\omega_r)^2}} \tag{6.75}$$

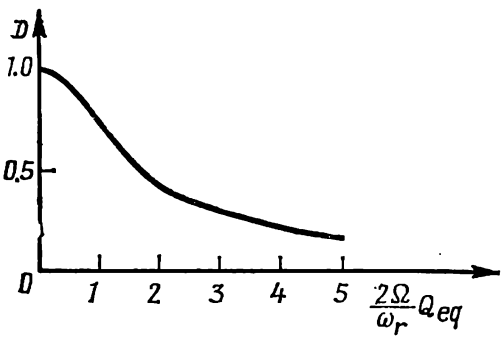


Fig. 6.18. Dependance of the demodulation coefficient in a resonance amplifier on the modulation frequency

The curve of D as a function of Ω, shown in Fig. 6.18, corresponds to the right-hand branch of the resonance curve of the tank circuit.

2. The envelope of the output oscillation lags in phase behind the envelope of the input oscillation by an angle

$$\xi_0 = \arctan a_{eq} = \arctan(2\Omega Q_{eq}/\omega_r) \tag{6.76}$$

The above results for steady-state tone modulation can also be easily obtained by analyzing the passage through the circuit of the individual spectral components of the modulated wave.

By writing expression (6.64) in the form

$$a(t) = E_0 \cos(\omega_0 t + \theta_0) + (ME_0/2) \cos[(\omega_0 + \Omega)t + \theta_0 + \gamma_0] + (ME_0/2) \cos[(\omega_0 - \Omega)t + \theta_0 - \gamma_0] \quad (6.77)$$

one can easily set up a similar expression for the output voltage of the amplifier.

Taking into account that the transfer function of the amplifier for frequencies ω_0, $\omega_0 + \Omega$, and $\omega_0 + - \Omega$ is respectively equal to [see formula (6.45)]

$$\mathrm{K}(i\omega_0) = \mathrm{K}_1(0) = -K_{\max}$$

$$\mathrm{K}[i(\omega_0 + \Omega)] = \mathrm{K}_1(i\Omega) = -\frac{K_{\max}}{1 + i\Omega\tau_t} = -\frac{K_{\max}}{\sqrt{1 + \Omega^2\tau_t^2}} e^{-i\xi_0}$$

$$\mathrm{K}[i(\omega_0 - \Omega) = \mathrm{K}_1(-i\Omega) = -\frac{K_{\max}}{1 - i\Omega\tau_t} = -\frac{K_{\max}}{\sqrt{1 + \Omega^2\tau_t^2}} e^{i\xi_0}$$

we may write

$$a_{out}(t) = -K_{\max}E_0 \left\{ \cos(\omega_0 t + \theta_0) + \frac{M}{2} \frac{1}{\sqrt{1 + \Omega^2\tau_t^2}} \times \cos[(\omega_0 + \Omega)t + \theta_0 + \gamma_0 - \xi_0] + \frac{M}{2} \frac{1}{\sqrt{1 + \Omega^2\tau_t^2}} \cos[(\omega_0 - \Omega)t + \theta_0 - \gamma_0 + \xi_0] \right\}$$

Reducing this expression yields expression (6.74).

The meaning of this result is explained in Fig. 6.19*a* which illustrates the position of the spectrum of the input oscillation with respect to the resonance characteristic of the tank circuit. The higher the modulation frequency Ω, the greater the relative attenuation of the amplitudes of side-frequency oscillations and, therefore, the shallower the depth of modulation.

The results obtained from the consideration of tone modulation enable one to get an idea of the general picture of phenomena that occur when transmitting oscillations modulated in amplitude by a *complex message*. Such a message contains oscillations of different frequencies Ω, which are attenuated to a different degree: the higher the frequency, the higher the demodulation. Since the output voltage of the detector of a receiver is proportional to the modulation factor, there takes place a relative attenuation of the upper-frequency oscillations of the message. Thus, the function $D(\Omega)$ defines the *degree of linear frequency distortion of the message being transmitted.*

The message is also subjected to a delay. This is explained by the fact that the phase shift of the envelope (in tone modulation) depends on frequency. The tank circuit affects the message contained in the envelope in the same manner as a low-pass filter, should the message be passed through it.

The magnitude of delay is defined by the slope of the phase characteristic:

$$t_0 = \left|\frac{d\varphi}{d\Omega}\right| = \left|\frac{d\left(\text{arc tan}\,\frac{2\Omega}{\omega_r}Q_{eq}\right)}{d\Omega}\right| = \frac{1}{1+\left(\frac{2\Omega}{\omega_r}Q_{eq}\right)^2}\frac{2Q_{eq}}{\omega_r} \qquad (6.78)$$

The delay is usually determined by the slope of the phase characteristic at the point $\Omega = 0$. In this case, we have

$$t_0 = 2Q_{eq}/\omega_r = \tau_t \qquad (6.79)$$

So, *the delay of a message in a single-tuned circuit, whose passband is wide enough for statisfactory transmission of the message spectrum, is equal to the time constant of the circuit.*

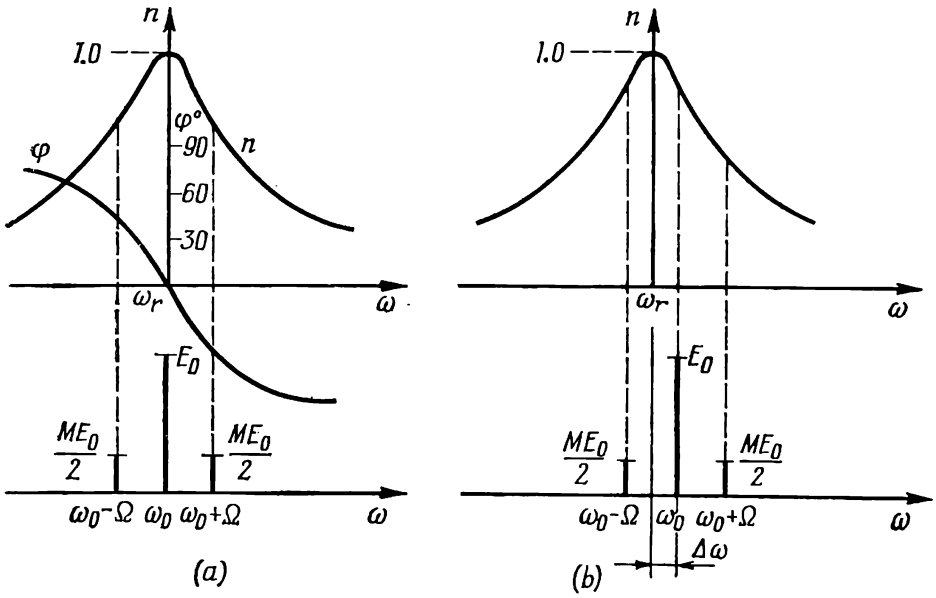


Fig. 6.19. Position of the spectrum of the modulated oscillation with respect to the frequency characteristic curve of an amplifier
(*a*) at accurate tuning; (*b*) at detuning

Now, let us consider the case of inexact tuning of the tank circuit to the carrier frequency of the modulated wave (Fig. 6.19*b*). The difference between frequencies ω_0 and ω_r results in the *asymmetry of the side-frequency oscillations at the amplifier output.* The development of the asymmetry is illustrated by the vector diagram of the output voltage shown in Fig. 6.20.

In this diagram, the vector *OD* represents the carrier whose phase lags behind the phase (assumed to be zero) of the input e.m.f. by an angle θ_0 (since Fig. 6.19*b* corresponds to a positive detuning $\Delta\omega =$

$= \omega_0 - \omega_r > 0$). In this case, the amplitude of the upper side-frequency oscillation (vector DC_1) is much lower than that of the lower side-frequency oscillation (vector DC_2). The length of the resultant vector OF representing the resultant oscillation varies according to a complex law that *does not coincide with the harmonic law governing the variation of the envelope of the input e.m.f.*

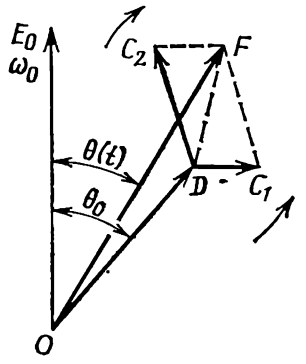


Fig. 6.20. Parasitic phase modulation in the case of asymmetry of the amplitudes of the side-frequency oscillations

It should be borne in mind that to restore the message being transmitted at the output of a radio channel employing amplitude modulation, use is made of an amplitude detector which is a nonlinear device. The output voltage of the detector is proportional to the *envelope* of the modulated wave. From this it follows that any disturbance in the symmetry of the amplitudes and phases of the side-frequency oscillations due to inexact tuning of the tank circuit to the carrier frequency ω_0 results in a *nonlinear distortion* of the transmitted message. This distortion manifests itself in the development of oscillations of new frequencies that are multiples of the useful modulation frequency.

Besides the distortion of the shape of the envelope, there also occurs a parasitic phase modulation of the oscillation, because during the rotation of the vectors DC_1 and DC_2 (Fig. 6.20) the phase θ (t) of the vector OF relative to the phase (taken as reference) of the carrier oscillation of the e.m.f. varies continuously. In some cases, this may result in an additional distortion of the signal.

The results obtained above can easily by applied to any oscillatory circuit, e.g., to multiple-tuned circuits. If the resonance curve of such a circuit is symmetrical with respect to the carrier frequency ω_0, the right-hand branch of this curve may be regarded as a characteristic of the coefficient D (see Fig. 6.18).

6.10. TRANSMISSION OF A PHASE-SHIFT KEYED OSCILLATION THROUGH A RESONANCE CIRCUIT

Along with continuous phase modulation, *phase-shift keying* also finds application in radio engineering. This process consists in changing suddenly the phase of a high-frequency oscillation by 180° at definite instants of time (Fig. 6.21*a*), the amplitude and frequency of the oscillation being kept constant. In Fig. 6.21*b*, the phases 0 and π alternate periodically, but when transmitting real signals, the law of phase alternation may be more complex.

Let us discuss the phenomena in resonance circuits, that occur at the moments the phase of the input signal is changed stepwise. Let us assume that the intervals T_1 between two adjacent phase steps are much longer than the duration of the transients arising in the circuit, so that each of the step scan be analyzed separately from the previous ones.

To reveal the principal aspect of the problem, let us restrict ourselves to a simple case—that of transmission of a phase-shift keyed

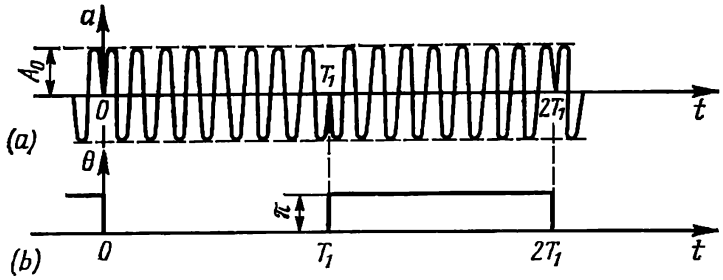


Fig. 6.21. (*a*) Phase-shift-keyed oscillation and (*b*) the character of the phase change

signal through a solitary tank circuit tuned to the signal frequency ω_0, i.e., $\omega_0 = \omega_r$.

Let the time origin coincide with the moment the phase is changed, as shown in Fig. 6.21. Then, for $t > 0$, the output signal, in accordance with the superposition principle, can be represented as the sum of a free oscillation remaining after the previous signal has ceased and a rising oscillation caused by the action of the next signal whose carrier phase differs from that of the previous signal by 180°.

Ignoring the difference between the natural frequency ω_{fr} of the tank circuit and the resonance frequency ω_r, we may write the following expressions for the two oscillations mentioned above:

$$\left.\begin{aligned} a_1(t) &= A_0 e^{-\alpha_t t}\cos\omega_r t, \\ a_2(t) &= -A_0(1-e^{-\alpha_t t})\cos\omega_r t \end{aligned}\right\} \qquad (6.80)$$

The minus sign on the right-hand side of the second expression accounts for the phase reversal.

The resultant signal at the output of the circuit (Fig. 6.22) is

$$\begin{aligned} s_{out}(t) = a_1(t) + a_2(t) &= (-A_0 + A_0 e^{-\alpha_t t} + A_0 e^{-\alpha_t t})\cos\omega_r t \\ &= -A_0(1-2e^{-\alpha_t t})\cos\omega_r t \end{aligned} \qquad (6.81)$$

Because of the time-lag nature of the circuit, the stepwise change of the input signal phase results in a change of the output signal ampli-

tude. At an instant $t_0 = 0.69/\alpha_t$, when $e^{-\alpha_t t_0} = 1/2$, the envelope vanishes. The lower the α_t (or the higher the quality factor of the circuit), the longer the t_0, i.e., the longer the time required for an oscillation with a new phase to reach its steady state.

In more complex oscillatory circuits, and also where there is a difference between the frequencies ω_0 and ω_r, the process becomes

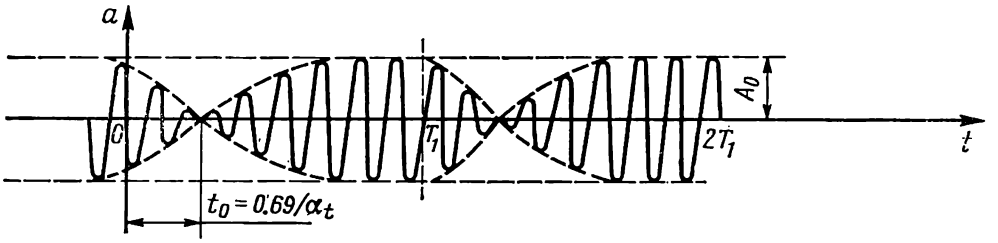


Fig. 6.22. Parasitic amplitude modulation in a resonance circuit due to a sudden change in the phase of the input e.m.f.

more involved: the development of a parasitic envelope fluctuation is accompanied by a disturbance of the nature of the phase change. Instead of changing stepwise, the phase smoothly changes from its former value to the new one. In this case, the method of determining the structure of the output signal remains the same, but $a_1(t)$ and $a_2(t)$ in the expression for $s_{out}(t)$ are now oscillations of different frequencies. After calculating the modulus and argument of the resultant oscillation, one can easily find the envelope and phase of the output signal.

6.11. TRANSMISSION OF A FREQUENCY-SHIFT KEYED OSCILLATION THROUGH A SELECTIVE CIRCUIT

Assume that a signal at the input of a selective circuit has a waveform shown in Fig. 6.23*a*. At certain instants of time the frequency changes stepwise from ω_1 to ω_2 and back from ω_2 to ω_1, the amplitude remaining constant and the phase having no discontinuities at the moments the frequency is changed. The last assumption is dictated by the desire to reveal the effect on the output signal parameters of frequency-shift keying alone, i.e., without considering the influence of phase-shift keying (discussed in the preceding section).

Assume that the moment $t = 0$ coincides with the moment the frequency changes from ω_1 to ω_2 (Fig. 6.23*b*) and that by $t = 0$ all the transients associated with the preceding frequency change have already ceased. Thus, for $t < 0$, the output signal is a harmonic oscillation with a frequency ω_1 and a constant amplitude A_0.

At first sight it might seem that a stepwise change in the frequency of the input signal, occurring without any changes in the amplitude

or phase, would cause no transients. However, this is not the case, since any frequency change in an energy-storing system is inevitably accompanied by a change in the stored energy.

The main idea upon which the following discussion is based, is that an instantaneous change in the frequency of the input e.m.f. is equivalent to switching off the old e.m.f. of frequency ω_1 and switching on at the same moment a new e.m.f. of frequency ω_2. A similar method was used in Sec. 6.10 for a stepwise change of the phase of the input signal, but here the problem is somewhat complicated because of the difference in frequency between different addends.

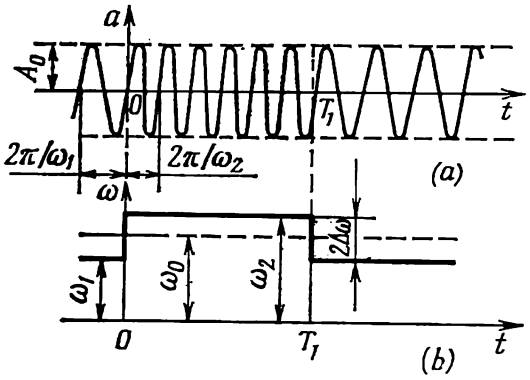


Fig. 6.23. (*a*) Frequency-shift-keyed oscillation and (*b*) the character of the frequency variation

Thus, for $t > 0$, the resultant oscillation at the output of a linear circuit is

$$a_{out}(t) = a_1(t) + a_2(t) \tag{6.82}$$

where $a_1(t)$ is a free oscillation due to the switching off at $t = 0$ of the old e.m.f. (of frequency ω_1) and $a_2(t)$ is a rising oscillation caused by the switching on of the new e.m.f. (of frequency ω_2).

Consider a solitary tank circuit with the output voltage taken from across the capacitor (Fig. 6.24). Let us assume that the reso-

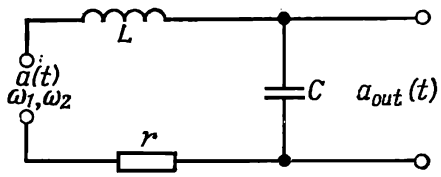


Fig. 6.24. Tank circuit excited by a frequency-shift-keyed oscillation

nance frequency of the circuit (ω_r) is equal to ω_0 and that the frequency shift $2\Delta\omega$ (see Fig. 6.23*b*) is symmetrical with respect to ω_0:

$$\omega_1 = \omega_0 - \Delta\omega = \omega_r - \Delta\omega, \quad \omega_2 = \omega_0 + \Delta\omega = \omega_r + \Delta\omega$$

Then, in terms of the designations adopted in Sec. 6.8 and in accordance with the second addend in expression (6.58) where the constant $K_{\max} E_0$ is replaced by Q, the free oscillation may be represented as

follows:

$$a_1(t) = \frac{Q}{\sqrt{1+\Delta\omega^2\tau_t^2}} e^{-t/\tau_t} \sin(\omega_r t + \theta_0 - \varphi_1) \qquad (6.83)$$

It should be noted that the free oscillation is taken here with a plus sign since it results from the switching off of the old e.m.f. and not from the switching on of a new one. The cosine is replaced by sine, because the voltage is taken from across the capacitor that is connected into a series-resonance circuit. Moreover, it should be borne in mind that for the frequency ω_1, which is lower than the resonance frequency of the circuit, the current in the circuit leads the e.m.f. so that the angle φ_1 is negative, i.e.,

$$\varphi_1 = \text{arc tan}\,(-\Delta\omega\tau_t) = -\text{arc tan}\,\Delta\omega\tau_t$$

Thus, denoting arc tan $\Delta\omega\tau_t = \varphi$, we obtain

$$a_1(t) = \frac{Q}{\sqrt{1+\Delta\omega^2\tau_t^2}} e^{-t/\tau_t} \sin(\omega_r t + \theta_0 + \varphi)$$

$$= \frac{Q}{\sqrt{1+\Delta\omega^2\tau_t^2}} e^{-t/\tau_t} [\cos\varphi \sin(\omega_r t + \theta_0) + \sin\varphi \cos(\omega_r t + \theta_0)] \qquad (6.84)$$

To determine $a_2(t)$, one may use an expression similar to (6.58), in which the constant $K_{\max} E_0$ should be replaced by Q, the frequency ω_0, by the frequency ω_2 of the new e.m.f., and the cosines, by sines, and the angle φ should be defined by the expression $\varphi =$ $=$ arc tan $\Delta\omega\tau_t$, as in (6.84).

So,

$$a_2(t) = \frac{Q}{\sqrt{1+\Delta\omega^2\tau_t^2}} [\sin(\omega_2 t + \theta_0 - \varphi) - e^{-t/\tau_t} \sin(\omega_r t + \theta_0 - \varphi)]$$

After substituting $\omega_2 = \omega_r + \Delta\omega$, this expression is reduced to the form

$$a_2(t) = \frac{Q}{\sqrt{1+\Delta\omega^2\tau_t^2}} \{\cos(\Delta\omega t - \varphi) \sin(\omega_r t + \theta_0) + \sin(\Delta\omega t - \varphi)$$

$$\times \cos(\omega_r t + \theta_0) - e^{-t/\tau_t} [\cos\varphi \sin(\omega_r t + \theta_0) - \sin\varphi \cos(\omega_r t + \theta_0)]\} \qquad (6.85)$$

Adding together expressions (6.84) and (6.85), we get

$$a_{out}(t) = \frac{Q}{\sqrt{1+\Delta\omega^2\tau_t^2}} \{\cos(\Delta\omega t - \varphi) \sin(\omega_r t + \theta_0) + [\sin(\Delta\omega t - \varphi)$$

$$+ 2e^{-t/\tau_t} \sin\varphi] \cos(\omega_r t + \theta_0)\} = A_{out}(t) \sin[\omega_r t + \theta_0 + \xi(t)] \qquad (6.86)$$

The envelope $A_{out}(t)$ and the variable part $\xi(t)$ of the phase of the output signal are determined by the following expressions:

$$A_{out}(t)=\frac{Q}{\sqrt{1+\Delta\omega^2\tau_t^2}}\sqrt{1+4e^{-t/\tau_t}\sin\varphi\sin(\Delta\omega t-\varphi)+4e^{-2t/\tau_t}\sin^2\varphi} \tag{6.87}$$

$$\xi(t)=\arctan\frac{\sin(\Delta\omega t-\varphi)+2e^{-t/\tau_t}\sin\varphi}{\cos(\Delta\omega t-\varphi)} \tag{6.88}$$

In this case, of principal interest is the law governing the variation of the output oscillation frequency:

$$\omega(t)=\omega_r+\frac{d\xi(t)}{dt}=\omega_r+\Delta\omega(t)$$

After differentiation we find that

$$\Delta\omega(t)=\Delta\omega \times \frac{1-2e^{-\alpha_t t}\sin\varphi\left[\frac{\alpha_t}{\Delta\omega}\cos(\Delta\omega t-\varphi)-\sin(\Delta\omega t-\varphi)\right]}{1+4e^{-\alpha_t t}\sin\varphi\left[\sin(\Delta\omega t-\varphi)+e^{-\alpha_t t}\sin\varphi\right]} \tag{6.89}$$

where $\alpha_t=1/\tau_t$.

Expression (6.89) can be simplified, because

$$\alpha_t/\Delta\omega=1/\Delta\omega\tau_t=\cot\varphi$$

and the numerator is reduced to the form

$$1-2e^{-\alpha_t t}\sin\varphi\left[\frac{\cos\varphi}{\sin\varphi}\cos(\Delta\omega t-\varphi)-\sin(\Delta\omega t-\varphi)\right]$$
$$=1-2e^{-\alpha_t t}\cos\Delta\omega t$$

Thus, finally, the relative detuning is

$$Y=\frac{\Delta\omega(t)}{\Delta\omega}=\frac{1-2e^{-\alpha_t t}\cos\Delta\omega t}{4+4e^{-\alpha_t t}\sin\varphi[\sin(\Delta\omega t-\varphi)+e^{-\alpha_t t}\sin\varphi]}$$
$$=\frac{1-2e^{-\Delta\omega t/b}\cos\Delta\omega t}{1+4e^{-\Delta\omega t/b}\sin\varphi\left[\sin(\Delta\omega t-\varphi)+e^{-\Delta\omega t/b}\sin\varphi\right]} \tag{6.90}$$

where $b=\Delta\omega/\alpha_t$.

The graphs $Y(\Delta\omega t)$ for several values of the parameter b are plotted in Fig. 6.25. Note that the passband of the circuit, determined by the attenuation of the signal to $1/\sqrt{2}$ of the maximum value is equal to $2\alpha_t=\omega_r/Q$. Therefore, the parameter b is none other than the ratio of the total signal frequency shift $2\Delta\omega$ to passband $2\alpha_t$.

In Fig. 6.26 the independent variable is $\alpha_t t$, hence equation (6.90) assumes the form

$$Y=\frac{1-2e^{-\alpha_t t}\cos(b\alpha_t t)}{1+4e^{-\alpha_t t}\sin\varphi\left[\sin(b\alpha_t t-\varphi)+e^{-\alpha_t t}\sin\varphi\right]} \tag{6.91}$$

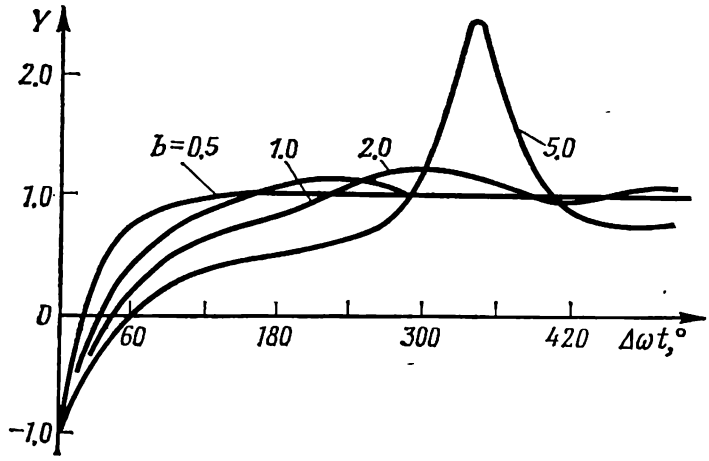


Fig. 6.25. Setting in of the oscillation frequency in a tuned circuit in the case of a stepwise change of the excitation frequency, depending on the parameter $b = \Delta\omega/\alpha_t$

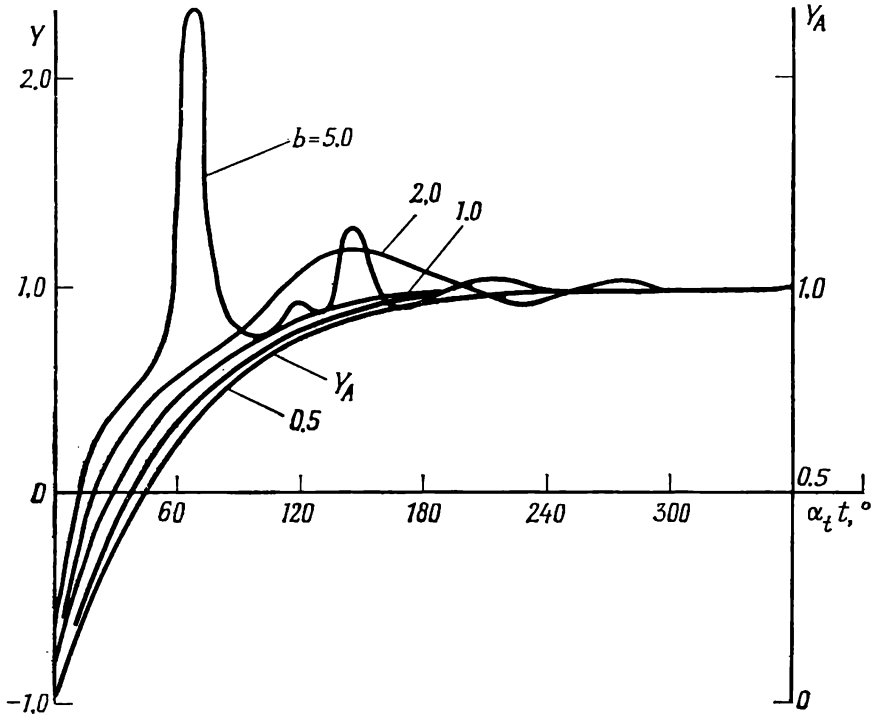


Fig. 6.26. The same as in Fig. 6.25 but for dimensionless time $\alpha_t t$

It should be noted that $\sin \varphi = b/\sqrt{1+b^2}$. Shown in the same figure is also the curve $Y_A = 1 - e^{-\alpha_t/t}$ that illustrates the manner in which the amplitude of current approaches its steady state when applying to the tank circuit an e.m.f. with a frequency ω_0 equal to the resonance frequency of the circuit.

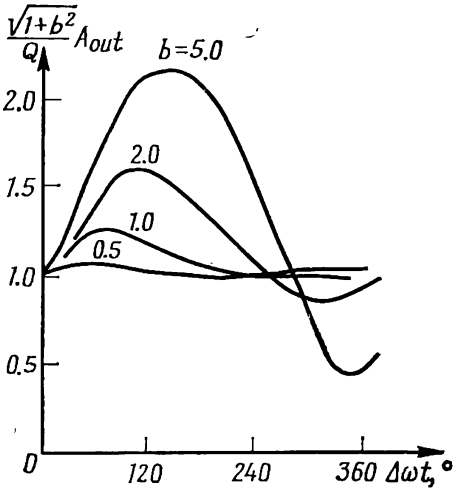


Fig. 6.27. Amplitude variation accompanying frequency-shift keying

From Fig. 6.26 it is clear that for $b \leqslant 0.5$, i.e., for $\Delta\omega/\alpha_t \leqslant 0.5$, the manner in which frequency reaches its steady state practically does not differ from that in which the amplitude of current reaches its steady-state value when the e.m.f. is suddenly switched on.

For $b > 0.5$, the discrepancy between the curves of Y and Y_A becomes noticeable. To complete the picture, let us consider the variation of the output signal amplitude.

Substituting $\Delta\omega\tau_t = \Delta\omega/\alpha_t = b$, $\sin\varphi = b/\sqrt{1+b^2}$, and $\cot\varphi = 1/b$ into expression (6.87) we get

$$A_{out}(t) = \frac{Q}{\sqrt{1+b^2}} \times \sqrt{1 + \frac{4b^2}{1+b^2} e^{-\Delta\omega t/b}\left(\frac{1}{b}\sin\Delta\omega t - \cos\Delta\omega t\right) + \frac{4b^2 e^{-2\Delta\omega t/b}}{1+b^2}} \qquad (6.92)$$

The graphs of the function $(\sqrt{1+b^2}/Q)\, A_{out}(\Delta\omega t)$, plotted for some values of the parameter b (Fig. 6.27), show that for $b \geqslant 0.5$ frequency variation is accompanied by substantial amplitude variations. When $b = 1$, i.e., when the frequency deviation reaches the boundaries of the passband of the circuit, the amplitude variation comes to about 25% of the steady-state value.

The inequality $b = \Delta\omega/\alpha_t \leqslant 0.5$ may be taken as a condition for monotonous rise of $\Delta\omega\,(t)$ without any significant amplitude variations. The duration of the interval T_1 (Fig. 6.23) must be sufficiently long compared to the time constant $\tau_t = 1/\alpha_t$ of the tank circuit. We may see that the last requirement is the same as in the case of transmission of radio pulses having a square envelope.

6.12. TRANSMISSION OF A FREQUENCY-MODULATED OSCILLATION THROUGH A SELECTIVE CIRCUIT

In Sec. 6.9 it was shown that with harmonic amplitude modulation, the transmission of oscillations through a circuit which is exactly tuned to the carrier frequency is not accompanied by any change in the shape of the envelope, and only the depth of modulation is decreased.

In the case of frequency modulation, the nonuniformity of the amplitude-frequency characteristic and the nonlinearity of the phase-frequency characteristic affect the parameters of the output oscillation in a more complicated manner. Even in the case of harmonic frequency modulation, the spectrum of the oscillation usually contains a large number of pairs of side frequencies. A disturbance of the normal amplitude and phase relations between individual pairs of side-frequency oscillations leads to a distortion of the modulating function, even if the circuit characteristics are fully symmetrical with respect to the carrier frequency.

In frequency modulation, circuit effects may manifest themselves in

(a) a distortion of the law governing the variation of the instantaneous frequency and instantaneous phase of the oscillation;

(b) a change in the amplitude of the useful frequency deviation, depending on the modulation frequency Ω;

(c) the development of a parasitic amplitude modulation.

The output voltage of a receiver with a frequency detector is proportional to the variation of the instantaneous frequency of the oscillation being detected. Therefore, any distortion of the law governing the variation of the instantaneous frequency in the tuned circuits of the transmitter and receiver results in *nonlinear distortion of the signal*, which manifests itself at the detector output in the form of additional voltages whose frequencies are multiples of the modulation frequency Ω.

The second of the above-mentioned changes in the parameters of a frequency-modulated oscillation leads to nonuniformity of the frequency characteristic of an FM radio channel and hence, to *frequency* (linear) distortion of the signal.

Let us consider the action on a resonance circuit of an e.m.f. whose frequency changes according to

$$\omega(t) = \omega_0 + \omega_d \cos \Omega t \qquad (6.93)$$

The amplitude of the e.m.f. is considered to be strictly constant so that the e.m.f. can be represented by the expression [see Sec. (3.23)]

$$e(t) = E_0 \cos(\omega_0 t + m \sin \Omega t)$$

Let us denote the complex transfer function of the circuit as

$$\mathrm{K}(i\omega) = K(\omega)\, e^{i\varphi(\omega)}$$

The modulus $K(\omega)$ and phase $\varphi(\omega)$ for an ordinary resonance circuit are shown in an approximate form in Fig. 6.28*a*. Since $\varphi(\omega)$ is taken with a plus sign, the phase characteristic $\varphi(\omega)$ has a negative slope within the passband of the circuit*. The frequency spectrum and the variation of the instantaneous frequency $\omega(t)$ of the input e.m.f. are shown in Fig. 6.28*b* and *c*. Oscillatory circuits are usually tuned to the centre frequency of the modulated wave, therefore, Fig. 6.28 and further discussion refer to the case $\omega_r = \omega_0$.

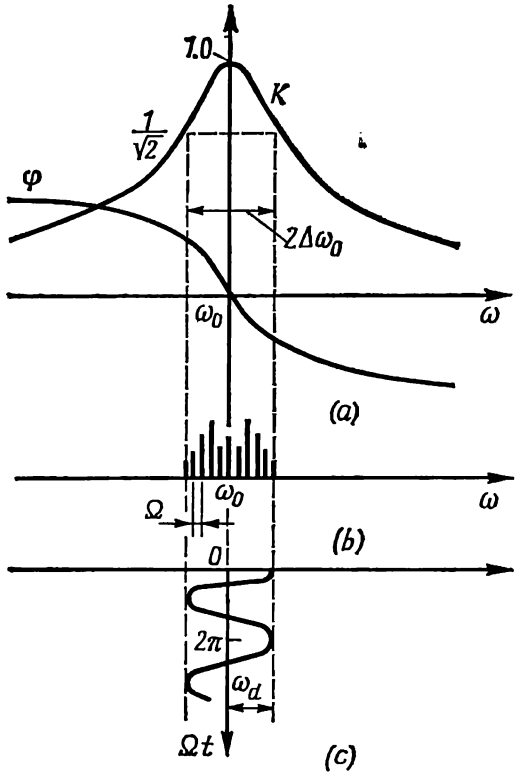


Fig. 6.28. (*a*) Transfer function of a circuit, (*b*) spectrum of FM oscillation, and (*c*) graph of the instantaneous frequency of this oscillation

To find the output oscillation of the circuit, one may, in principle, use the same method as in the case of amplitude modulation (see Sec. 6.9). In so doing, it is necessary to take into account amplitude and phase changes for each pair of the side-frequency components of the e.m.f. in accordance with the curves of $K(\omega)$ and $\varphi(\omega)$. But such a method, although very accurate, is suitable for very low modulation indices only, i.e., if the spectral composition of an FM oscillation differs but slightly from that of an AM oscillation. In practice, however, one most often encounters modulation in which the number of components in the frequency band in use is so large that the application of the spectral method would entail very great and sometimes insurmountable calculation difficulties. In such cases, one has to resort to approximate methods permitting the oscillation at the output of a circuit to be determined proceeding from the given modulating function and frequency-phase characteristics of the circuit without resolving the input e.m.f. into a spectrum.

These methods, called *instantaneous frequency methods*, are based on the assumption that frequency varies *slowly*. The modulation frequency is assumed to be so low that at any instant of time the

* Phase shifts that are independent of frequency (e.g., the 90-degree shift in the circuit shown in Fig. 6.27) are not taken into account here.

amplitude and phase of the oscillation at the output of the circuit can be determined sufficiently accurately on the basis of the frequency and phase characteristics of the circuit in the same manner as under steady-state conditions. Thus, it is assumed that the output oscillation reaches its steady state almost simultaneously with the change in the frequency at the input of the circuit.

The longer the modulation period $2\pi/\Omega$ and the smaller time constant τ of the circuit, the nearer these assumptions to the truth. Since τ is inversely proportional to the passband $2\Delta\omega_0$ of the circuit, one of the conditions for the applicability of the instantaneous frequency method is the inequality $\Omega/\Delta\omega_0 \ll 1$.

The frequency Ω being the same, the rate of change of the instantaneous frequency depends on the amplitude of the frequency deviation ω_d, therefore, the fulfilment of the above condition alone is not sufficient. Some limitations must also be imposed on the ratio $\omega_d/\Delta\omega_0$. A more thorough analysis [5] shows that if $\omega_d/\Delta\omega_0$, is less than unity, or is close to it, the instantaneous frequency method proves accurate enough for all practical purposes.

With all the above conditions being satisfied, the voltage at the output of the circuit can be defined by the following expression:

$$v_{out}(t) = E_0 \operatorname{Re}[e^{i\psi(t)} K(i\omega)] = E_0 K(\omega) \operatorname{Re}\{e^{i[\psi(t)+\varphi(\omega)]}\}$$

where $\psi(t) = \omega_0 t + m \sin \Omega t$ is the phase of the e.m.f. at the input of the circuit (see Sec. 3.4) and $\varphi(\omega)$ is the argument of the transfer function of the circuit.

From this expression it is evident that the amplitude of the output voltage changes according to

$$V_{out}(t) = E_0 K(\omega) = E_0 K(\omega_0 + \omega_d \cos \Omega t)$$

and the instantaneous frequency, according to

$$\omega_{out}(t) = \frac{d\psi}{dt} + \frac{d\varphi}{dt}$$

Since the first term on the right-hand side of this expression represents the instantaneous frequency $\omega(t)$ of the input e.m.f., the quantity $\xi(t) = d\varphi/dt$ characterizes the effect of the circuit in question on the frequency of the output oscillation. With the above-mentioned condition for slow modulation being satisfied, ξ is, as a rule, small as compared with ω_d.

So,

$$\omega_{out}(t) = \omega(t) + \xi(t) \tag{6.94}$$

If the equation of the phase characteristic $\varphi(\omega)$ is known, then, substituting in accordance with (6.93) $\omega(t) = \omega_0 + \omega_d \cos \Omega t$ into (6.94) and differentiating with respect to t, we obtain a general expression for $\xi(t)$:

$$\xi(t) = \frac{d}{dt}[\varphi(\omega_0 + \omega_d \cos \Omega t)] \tag{6.95}$$

With periodic frequency modulation, $\xi(t)$ is also a periodic function of time, so it can be expanded into a Fourier series. When the circuit is tuned to the centre frequency ω_0, the phase characteristic is, as a rule, antisymmetric with respect to ω_0, therefore the Fourier series contains only odd harmonics: Ω, 3Ω, 5Ω, etc. Finally, when frequency varies according to (6.93), the derivative of φ, i.e., $\xi(t)$ is an odd function of time, hence, the Fourier series contains only sine terms. So,

$$\xi(t) = \mathscr{E}_1 \sin \Omega t + \mathscr{E}_3 \sin 3\Omega t + \ldots \tag{6.96}$$

where $\mathscr{E}_1$, $\mathscr{E}_3$, ... are the amplitudes of the harmonics of the function $\xi(t)$.

Substituting expression (6.96) into (6.94), we obtain

$$\omega_{out}(t) \approx \omega_0 + \omega_d \cos \Omega t + \mathscr{E}_1 \sin \Omega t + \mathscr{E}_3 \sin 3\Omega t + \ldots \approx \omega_0$$

$$+ \sqrt{\omega_d^2 + \mathscr{E}_1^2} \cos(\Omega t - \gamma) + \mathscr{E}_3 \sin 3\Omega t + \ldots \approx \omega_0 + \omega_d \cos(\Omega t + \gamma)$$

$$+ \mathscr{E}_3 \sin 3\Omega t + \ldots \tag{6.97}$$

The addend $\mathscr{E}_1^2$ under the radical sign is omitted as a quantity of a higher order of smallness compared to ω_d^2.

Comparison between expressions (6.93) and (6.97) allows one to draw a conclusion that the circuit affects the message so that the phase of the output oscillation lags behind that of the input signal by an angle γ determined by the expression

$$\gamma = \arctan(\mathscr{E}_1/\omega_d) \tag{6.98}$$

and odd harmonics appear in the modulating function. As mentioned above, the latter circumstance is usually of greatest importance.

Let us illustrate the application of the instantaneous frequency method with an example of a solitary tank circuit.

Implying by $\mathrm{K}(i\omega)$ the ratio of the complex amplitude of the voltage across the tank capacitor to the amplitude of the e.m.f. connected in series into the circuit, we obtain

$$\mathrm{K}(i\omega) = \frac{1/i\omega C}{r[1 + i(\omega - \omega_0)\tau_t]}$$

Taking into account that $\omega - \omega_0 = \omega_d \cos \Omega t$ and neglecting the variations of ω in the numerator (since ω_d is usually low compared to ω_0), we may write

$$\mathrm{K}(i\omega) \approx \frac{Q}{i(1 + i\omega_d \tau_t \cos \Omega t)} = \frac{Q}{\sqrt{1 + (\omega_d \tau_t \cos \Omega t)^2}} e^{i\varphi}$$

where

$$\varphi = -\left[\frac{\pi}{2} + \arctan(\omega_d \tau_t \cos \Omega t)\right]$$

On the basis of expression (6.95) we get

$$\xi(t) = \frac{d\varphi}{dt} = \frac{\Omega\omega_d\tau_t \sin \Omega t}{1+\omega_d^2\tau_t^2 \cos^2 \Omega t} \tag{6.99}$$

Using formula (2.24), we find

$$\mathscr{E}_1 = \frac{1}{\pi}\int_0^{2\pi} \xi(t) \sin \Omega t d(\Omega t), \qquad \mathscr{E}_3 = \frac{1}{\pi}\int_0^{2\pi} \xi(t) \sin 3\Omega t d(\Omega t)$$

After integrating [see (2.553.3), (2.554.2), and (3.644.3) in [7]], we obtain the following final formulas for the amplitudes of the

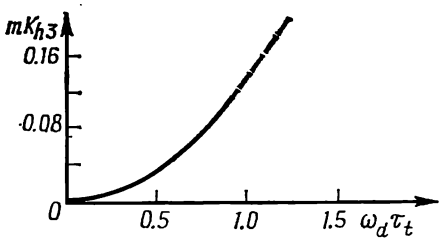


Fig. 6.29. Dependence of harmonic content on deviation ω_d at a specified value of the time constant of the tuned circuit

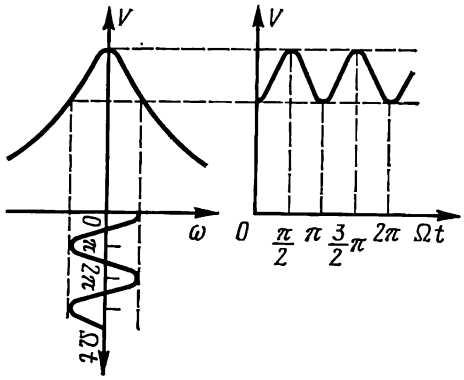


Fig. 6.30. Parasitic amplitude modulation in frequency modulation

first and third harmonics of the function $\xi(t)$:

$$\mathscr{E}_1 = \frac{2}{m\tau_t}(\sqrt{1+\omega_d^2\tau_t^2}-1), \quad \mathscr{E}_3 = \frac{2}{m\tau_t}\frac{(1-\sqrt{1+\omega_d^2\tau_t^2})^3}{\omega_d^2\tau_t^2} \tag{6.100}$$

where $m = \omega_d/\Omega$.

Then, using formula (6.98), we obtain the phase shift of the message

$$\gamma = \arctan\frac{\mathscr{E}_1}{\omega_d} = \arctan\left[\frac{2}{m\omega_d\tau_t}(\sqrt{1+\omega_d^2\tau_t^2}-1)\right] \tag{6.101}$$

Now one can easily determine the harmonic content in terms of the frequency 3Ω at the output of the frequency detector. For this purpose, one should divide the amplitude $\mathscr{E}_3$ of the third harmonic of the function ξ by the amplitude ω_d of the fundamental frequency Ω [see formula (6.100)]:

$$K_{h3} = \frac{\mathscr{E}_3}{\omega_d} = \frac{2}{m}\left|\frac{1-\sqrt{1+\omega_d^2\tau_d^2}}{\omega_d\tau_t}\right|^3 \tag{6.102}$$

The graph of the function $mK_{h3}(\omega_d\tau_t)$ is shown in Fig. 6.29. For $\omega_d\tau_t \ll 1$, formulas (6.101) and (6.102) are simplified:

$$\gamma \approx \Omega\tau_t,\ K_{h3} \approx (\omega_d\tau_t)^3/4m$$

As $\omega_d\tau_t \to 1$ (but $m \gg 1$), i.e., when the deviation is almost equal to the passband of the circuit, formulas (6.101) and (6.102) yield

$$\gamma = 0.8/m,\ K_{h3} = 0.13/m$$

Thus, where the instantaneous frequency method is applicable, the maximum distortion in a solitary tank circuit does not exceed a few fractions of a percent.

The amplitude variations of the output oscillation are easy to find. For this purpose, one may use the resonance curve of the circuit and make a graphic construction as shown in Fig. 6.30. One can easily see that the fundamental frequency of variation of the envelope of amplitudes V is twice the modulation frequency Ω.

Chapter 7

TRANSMISSION OF RANDOM OSCILLATIONS THROUGH LINEAR CIRCUITS WITH CONSTANT PARAMETERS

7.1. CONVERSION OF CHARACTERISTICS OF A RANDOM PROCESS

Restricting ourselves to the analysis of stationary random processes, let us set the following problem: applied to the input of a linear two-port network (Fig. 7.1) having a transfer function $\mathbf{K}(i\omega)$ and an impulse response $g(t)$ is a voltage $s(t)$ which is a random time function with an energy spectrum $W_s(\omega)$ and an autocorrelation function $B_s(\tau)$; it is necessary to find the energy spectrum $W_{s,out}(\omega)$.

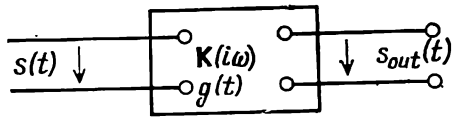


Fig. 7.1. Linear two-port network with constant parameters

and autocorrelation function $B_{s,out}(\omega)$ of the oscillation $s_{out}(t)$ at the output of the two-port network.

This problem can easily be solved by using the arguments advanced in Sec. 4.3 [see (4.31) through (4.33)] to define the energy spectrum of a random process.

If the spectral density $|X_{kT}(\omega)|$ of the kth realization in formula (4.32) is multiplied by the modulus $K(\omega)$ of the transfer function, one will obtain the spectral density of the same realization at the output of the two-port network. From this stems the following relation:

$$W_{s,out}(\omega) = W_s(\omega) K^2(\omega) \tag{7.1}$$

The squaring of the transfer function is explained by the fact that $W_s(\omega)$ is the spectral *power* density of the random function, while $K(\omega)$ defines the ratio between the *voltages* (or currents) at the output and input.

The phase-frequency characteristic of the filter is of no consequence in determining $W_{s,out}(\omega)$, because the phases of the spectral components of the input oscillation are random, so additional phase shifts occurring in the filter do not change a thing in the oscillation structure.

The autocorrelation function of the random process at the filter output is defined by expression (4.39):

$$B_{s,\,out}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W_{s,\,out}(\omega)\, e^{i\omega\tau}\, d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} W_s(\omega)\, K^2(\omega)\, e^{i\omega\tau}\, d\omega \tag{7.2}$$

The relations between the statistical characteristics of the random processes at the input and output of the circuit can also be derived using the impulse response of the circuit.

Indeed, since the spectral function $W_s(\omega)$ corresponds to the autocorrelation function

$$B_s(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W_s(\omega)\, e^{i\omega\tau}\, d\omega$$

and the spectral function $K^2(\omega)$ corresponds to

$$B_g(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} K^2(\omega)\, e^{i\omega\tau}\, d\omega$$

i.e., to the autocorrelation function of the impulse response $g(t)$ [see, for example, (2.118) in which $S^2(\omega)$ must be replaced by $K^2(\omega)$], the product of the spectral functions $W_s(\omega)$ and $K^2(\omega)$ corresponds to the convolution of the functions $B_s(\tau)$ and $B_g(\tau)$ [see (2.64)]:

$$B_{s,\,out}(\tau) = \int_{-\infty}^{\infty} B_s(x)\, B_g(\tau - x)\, dx \tag{7.3}$$

Thus, given the autocorrelation functions $B_s(\tau)$ and $B_g(\tau)$, one can determine the autocorrelation function $B_{s,\,out}(\tau)$ and then find the energy spectrum

$$W_{s,\,out}(\omega) = \int_{-\infty}^{\infty} B_{s,\,out}(\tau)\, e^{-i\omega\tau}\, d\tau \tag{7.4}$$

So, neither the spectral nor the correlation analysis of the transmission of a stationary random process through a linear circuit with constant parameters presents any difficulties.

It is different with defining the distribution law of a random process at the output of a linear circuit. In the general case where the distribution of the process at the input is random, finding the distribution at the output of a time-lag circuit is a fairly complex problem.

These difficulties disappear where the distribution of the input process is normal, because *no matter what linear operation* (amplifi-

cation, filtering, differentiation, integration, etc.) *a normal process is subjected to, the distribution of the process remains normal* and only the functions $B(\tau)$ and $W(\omega)$ are changed. Therefore, if the probability density of the input process (with zero mean)

$$p(s) = \frac{1}{\sqrt{2\pi}\,\sigma_s} \exp\left(-\frac{s^2}{2\sigma_s^2}\right)$$

is known, the probability density at the output of a linear circuit will be

$$p(s_{out}) = \frac{1}{\sqrt{2\pi}\,\sigma_{s,\,out}} \exp\left(-\frac{s_{out}^2}{2\sigma_{s,\,out}^2}\right) \tag{7.5}$$

In accordance with (4.36), the variance of the output process is defined by the expression

$$\sigma_{s,\,out}^2 = B_{s,\,out}(0) = \frac{1}{2\pi} \times \int_{-\infty}^{\infty} W_s(\omega)\, K^2(\omega)\, d\omega \tag{7.6}$$

The analysis of transmission of normal processes through linear circuits is essentially reduced to correlation (or spectral) analysis.

7.2. CHARACTERISTICS OF INTRINSIC NOISE IN RADIO CIRCUITS

When analysing the transmission of signals through radio circuits, the *intrinsic* (*internal*) *noise* of the circuit must be taken into account along with the unavoidable distortion of the shape of the signals. Being superimposed on the signal, this noise limits the information capacity of this signal. The problem of noise becomes especially critical when amplifying weak signals.

Radioelectronic systems have two basic noise sources: the discrete structure of current in amplifying elements (transistors, tubes, etc.) and thermal motion of free electrons in the circuit conductors.

Let us consider the first source, taking as an example the shot effect typical of the electron current in amplifier devices. This current is a combination of pulses, each pulse being caused by the transfer of one electronic charge. The total current, which is a sum of a very large number of overlapping pulses arranged randomly along the time axis, is a stationary, ergodic random process obeying the central limit theorem*. Therefore, the distribution of the electron

* The substance of this theorem is reduced to the following: the distribution of a sum of a large number of random, mutually independent terms, among which there are no dominating ones, is close to the normal regardless of the distributions of the individual terms.

current can be considered to be normal with a probability density

$$p(i) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{(i-I_0)^2}{2\sigma_i^2}\right] \tag{7.7}$$

The d-c current component I_0 and the average power σ_i^2 of the fluctuation component can be found by the following reasoning.

Let the average number of pulses per second be equal to k_1. Since each pulse transfers a single electronic charge e, the total amount of electricity transferred per second is, on the average, equal to $k_1 e$. This is the d-c current component. Thus we have

$$I_0 = k_1 e \tag{7.8}$$

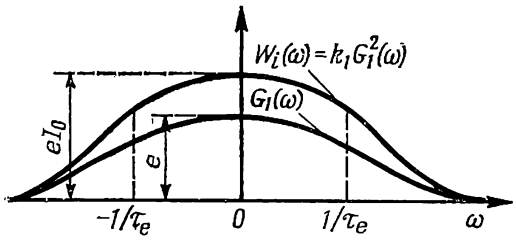


Fig. 7.2. Spectral density $G_1(\omega)$ of a solitary pulse and the energy spectrum $W_i(\omega)$ of a random process

Let us consider the spectral density $G_1(\omega)$ of a single current pulse $i_e(t - t_k)$ caused by the transfer of one electronic charge e (t_k is the moment of liberation of the electron). Regardless of the shape of the pulse, the value of $G_1(\omega)$ at $\omega = 0$ is equal to the pulse area (see Sec. 2.9-1):

$$G_1(0) = \int_{-\infty}^{\infty} i_e(t - t_k)\,dt = e \tag{7.9}$$

The duration τ_e of the pulse $i_e(t)$ depends on the geometry of the electron device, the electric field strength in the interelectrode spaces, etc. The width of the pulse spectrum can roughly be taken equal to $2/\tau_e$. Thus, the modulus of the spectral density of the pulse $i_e(t - t_k)$ can be represented in the form of a graph like the one in Fig. 7.2 where the maximum ordinate is approximately equal to e.

According to Parseval's formula, the energy of a single pulse is

$$E_1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_1^2(\omega)\,d\omega$$

while the total energy of k_1 pulses per second, i.e., the *average power of the process* (with a circuit resistance of 1 ohm) is

$$\overline{i^2(t)} = k_1 E_1 + I_0^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} k_1 G_1^2(\omega)\,d\omega + I_0^2 = \sigma_i^2 + I_0^2 \tag{7.10}$$

The first term on the right-hand side of (7.10) defines the power of the fluctuation component of the current. The second term in (7.10) defines the power of the d-c component I_0.

From the first term on the right-hand side of expression (7.10) it follows that the energy spectrum of the fluctuation component of the electron current coincides in shape with the spectral energy density $G_1^2(\omega)$ of the individual pulses forming the random process

$$W_i(\omega) = k_1 G_1^2(\omega) \tag{7.11}$$

The graph of $W_i(\omega)$ is shown in Fig. 7.2.

Taking into account that $k_1 = I_0/e$ and also the fact that within the frequency band of about $2/\tau_e$ equality (7.9) holds true, we get*

$$W_i(\omega) \approx eI_0, \quad 0 < |\omega| < 1/\tau_e \tag{7.12}$$

Expressions (7.7) and (7.12) define the main statistical characteristics of the fluctuation current due to shot effect.

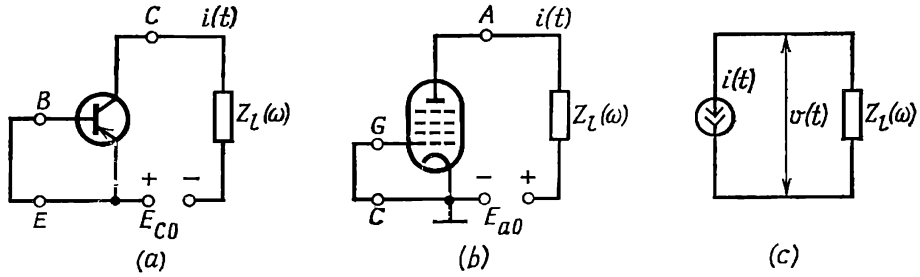


Fig. 7.3. (*a*) Transistor and (*b*) vacuum-tube amplifiers and (*c*) common equivalent circuit for fluctuation current

Now we can easily find the statistical characteristics of the noise voltage at the output of a circuit containing a "noisy" element. Figure 7.3*a* and *b* shows schematic diagrams of a transisitor and a vacuum-tube amplifier, while Fig. 7.3*c* shows a common equivalent circuit for the fluctuation current $i(t)$. The base-to-emitter (or, respectively, grid-to-cathode) input terminals are short-circuited to stress the absence of any external action on the amplifier. In the equivalent circuit, the noise source is represented by a current generator $i(t)$ whose statistical characteristics $p(i)$ and $W_i(\omega)$ are given above.

The noise voltage $v(t)$ across the linear load element $Z_l(\omega)$, as well as the current $i(t)$, obeys the normal distribution law

$$p(v) = \frac{1}{\sqrt{2\pi}\,\sigma_v} \exp\left(-\frac{v^2}{2\sigma_v^2}\right) \tag{7.13}$$

The energy spectrum of the voltage $v(t)$ is defined by the relation

$$W_v(\omega) = W_i(\omega)\, Z_l^2(\omega) \tag{7.14}$$

* The formula $W_i(\omega) = 2eI_0$ is also widely used in technical literature. In deriving this formula, the average power σ^2 is defined for positive frequencies only.

[compare with (7.1); in this case the impedance $Z_l(\omega)$ appears instead of the dimensionless transfer function $K(\omega)$].

Applying transformation (4.39) to (7.14), one can determine the autocorrelation function of the noise voltage at the amplifier output and also, the root-mean-square noise voltage σ_v.

Let us consider the mechanism of intrinsic noise formation in resistance-coupled and resonance amplifiers.

For the resistance-coupled amplifier shown in Fig. 5.14*a*, the impedance $Z_l(i\omega)$ is found by the formulas

$$Z_l(i\omega) = \frac{R\,(1/i\omega C_0)}{R+1/i\omega C_0}; \qquad Z_l^2(\omega) = \frac{R^2}{1+(\omega C_0 R)^2} \tag{7.15}$$

The time constant of the R-C_0 circuit is much greater than the pulse duration τ_e; correspondingly, the passband of the R-C_0 circuit, adjoining the zero frequency is many times narrower than the width of the energy spectrum $W_i(\omega)$ shown in Fig. 7.2. Therefore, the shot noise acting on the circuit may be regarded as white noise with an energy spectrum $W_i(\omega) = eI_0$. Then, according to formula (7.14),

$$W_v(\omega) = eI_0R^2/[1+(\omega C_0 R)^2] \tag{7.16}$$

and by formula (4.39)

$$B_v(\tau) = eI_0R^2\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{i\omega\tau}}{1+(\omega C_0R)^2}\,d\omega$$

$$= \frac{eI_0R^2}{(RC_0)^2}\frac{1}{\pi}\int_0^{\infty}\frac{\cos\omega\tau}{(1/RC_0)^2+\omega^2}\,d\omega$$

The integral on the right-hand side

$$\int_0^{\infty}\frac{\cos\omega\tau}{(1/RC_0)^2+\omega^2}\,d\omega = \frac{\pi}{2}RC_0\exp\left(-\frac{|\tau|}{RC_0}\right)$$

Thus,

$$B_v(\tau) = \frac{eI_0R}{2C_0}\exp\left(-\frac{|\tau|}{RC_0}\right) \tag{7.17}$$

At $\tau = 0$ this expression defines the variance σ_v^2 of the noise voltage and the root-mean-square noise voltage σ_v:

$$\sigma_v^2 = B_v(0) = \frac{eI_0R}{2C_0}; \qquad \sigma_v = \sqrt{\frac{eI_0R}{2C_0}} \tag{7.18}$$

The normalized autocorrelation function of noise assumes the form

$$R_v(\tau) = \exp(-|\tau|/RC_0) \tag{7.19}$$

The graphs of the energy spectrum $W_v(\omega)$ and of the function $R_v(\tau)$ are shown in Figs. 7.4 and 7.5.

In the given example, the correlation interval of the noise voltage is defined by the value of $|\tau|/RC_0 \approx 1$. The meaning of this result can easily be explained. The noise voltage across the load is generated by a set of random current pulses produced by individual electrons. Each of these pulses produces a voltage pulse whose duration is determined by the time constant of the load. When superimposing a large number of pulses, the relative rate of change of the total noise

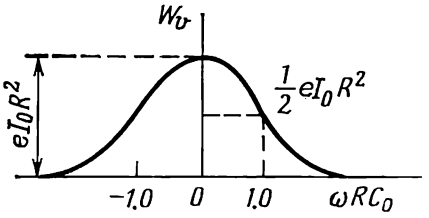


Fig. 7.4. Energy spectrum of the noise voltage at the output of a resistance-coupled amplifier

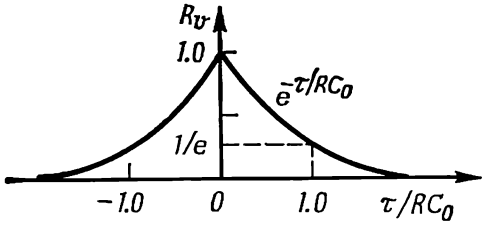


Fig. 7.5 .Normalized autocorrelation function of noise corresponding to the spectrum $W_v(\omega)$ (Fig. 7.4)

voltage $v(t)$ must be of the same order of magnitude as the rate of change of the individual pulses. Therefore, for the voltages taken at the instants t and $t + \tau$ to be independent, τ must not be shorter than the duration of the pulses forming the noise.

To evaluate quantitatively the level of the noise voltage due to shot effect, let us give the following example typical of an aperiodic amplifier: d-c current $I_0 = 10$ mA, load resistance $R = 5$ kΩ, capacitance $C_0 = 50$ pF.

Using formula (7.18), we find the root-mean-square noise voltage at the amplifier output:

$$\sigma_v \approx \sqrt{\frac{1.6\times10^{-19}\times10\times10^{-3}\times5\times10^{3}}{2\times50\times10^{-12}}} = 2.8\times10^{-4}\text{V} = 0.28\ \text{mV}$$

The voltage thus defined may be regarded as a result of application of some noise voltage to the amplifier input. With an amplification factor of K_a, the equivalent input noise voltage should be equated to the quantity $v_{rms} = \sigma_v/K_a$. It is this value that defines the lower threshold of the signal whose amplification by means of the given amplifier still makes sense.

In a similar way, we can consider the generation of noise in the tank circuit of a resonance amplifier whose circuit diagram is shown in Fig. 5.17.

By analogy with expression (7.14), let us find the energy spectrum

$$W_v(\omega) = W_i(\omega)\,Z_e^2(\omega) = eI_0Z_e^2(\omega) \qquad (7.20)$$

where

$$Z_e(i\omega) = \frac{Z_{e,r}}{1+ia_{eq}} \approx \frac{R_{sh}}{1+i\dfrac{2(\omega-\omega_r)}{\omega_r}Q_{eq}}$$

and $Z_{e,r} = R_{sh}$ is the impedance of the tank circuit (shunted by the resistor R_{sh}) at resonance. From this we find the square of the magni-

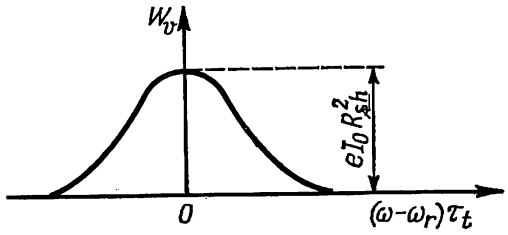


Fig. 7.6. Energy spectrum of the noise voltage at the output of a resonance amplifier

tude of the load impedance to be

$$Z_e^2(\omega) = R_{sh}^2/[1+(\omega-\omega_r)^2\tau_t^2] \tag{7.21}$$

where $\tau_t = 2Q_{eq}/\omega_r$ is the time constant of the tank circuit.

Thus,

$$W_v(\omega) = eI_0R_{sh}^2/[1+(\omega-\omega_r)^2\tau_t^2] \tag{7.22}$$

The graph of the energy spectrum $W_v(\omega)$ is shown in Fig. 7.6.

In this case, expression (4.39) for the autocorrelation function assumes the following form:

$$B_v(\tau) = eI_0R_{sh}^2\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\cos\omega\tau}{1+(\omega-\omega_r)^2\tau_t^2}d\omega$$

$$= eI_0R_{sh}^2\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\cos\omega\tau}{1+(\omega-\omega_r)^2\tau_t^2}d\omega$$

Transforming to a new variable $\omega_1 = \omega - \omega_r$, we obtain

$$B_v(\tau) = \frac{eI_0R_{sh}^2}{\pi}\int_{-\omega_r}^{\infty}\frac{\cos(\omega_1+\omega_r)\tau}{1+\omega_1^2\tau_t^2}d\omega_1$$

$$= \frac{eI_0R_{sh}^2}{\pi}\left[\cos\omega_r\tau\int_{-\omega_r}^{\infty}\frac{\cos\omega_1\tau}{1+\omega_1^2\tau_t^2}d\omega_1 - \sin\omega_r\tau\int_{-\omega_r}^{\infty}\frac{\sin\omega_1\tau}{1+\omega_1^2\tau_t^2}d\omega_1\right]$$

It should be noted that with a sufficiently high Q-factor of the tank circuit, the condition

$$\omega_r\tau_t = \omega_r(2Q_{eq}/\omega_r) = 2Q_{eq} \gg 1$$

is satisfied.

Therefore the lower limit $-\omega_r$ of the integrals can be replaced by $-\infty$. In this case, the second integral vanishes, because the integrand is odd with respect to the integration variable ω_1, and the first integral is reduced to

$$\int_{-\omega_r}^{\infty} \frac{\cos \omega_1 \tau}{1+\omega_1^2 \tau_t^2} d\omega_1 \approx \frac{2}{\tau_t^2} \int_0^{\infty} \frac{\cos \omega_1 \tau}{1/\tau_t^2+\omega_1^2} d\omega_1$$

because the integrand is even.

A similar integral was calculated when deriving formula (7.17). Using this result, we obtain

$$B_v(\tau) = \frac{eI_0 R_{sh}^2}{\pi} \cos \omega_r \tau \frac{2}{\tau_t^2} \frac{\pi \tau_t}{2} e^{-|\tau|/\tau_t} = \frac{eI_0 R_{sh}^2}{\tau_t} e^{-|\tau|/\tau_t} \cos \omega_r \tau$$

$$= eI_0 R_{sh}^2 \alpha_t e^{-\alpha_t |\tau|} \cos \omega_r \tau \qquad (7.23)$$

Here, $\alpha_t = 1/\tau_t$ denotes the attenuation of the tank circuit. Taking into account the fact that the attenuation factor of a tank circuit shunted by a resistor R_{sh} is $\alpha_t = 1/2R_{sh}C$, let us write formula (7.23) as follows:

$$B_v(\tau) = \frac{eI_0 R_{sh}}{2C} e^{-\alpha_t |\tau|} \cos \omega_r \tau \qquad (7.23')$$

From formulas (7.23) and (7.23′) it follows that, first, the mean square of the noise voltage across the tank circuit is

$$\sigma_v^2 = B_v(0) = eI_0 R_{sh}^2 \alpha_t = eI_0 R_{sh}/2C \qquad (7.24)$$

and hence, the root-mean-square noise voltage $\sigma_v = \sqrt{eI_0 R_{sh}/2C}$, and second, the normalized autocorrelation function is defined by the expression

$$R_v(\tau) = e^{-\alpha_t |\tau|} \cos \omega_r \tau = e^{-|\tau|/\tau_t} \cos \omega_r \tau \qquad (7.25)$$

The graph of the function $R_v(\tau)$ is shown in Fig. 7.7. In the case in question, the correlation interval is determined by the run of the envelope of the function $R_v(\tau)$, i.e., by the factor $e^{-|\tau|/\tau_t}$ in expression (7.25).

The noise voltage can be referenced to the input of a resonance amplifier, as in the case of an aperiodic amplifier, using the formula $v_{rms} = \sigma_v/K_a$ in which K_a implies the amplification factor at the resonance frequency.

The structure of the noise voltage across a high-Q tank circuit is shown in Fig. 4.17. The characteristics of a narrow-band random process, given in Sec. 4.6, fully apply to the shot noise in a resonance amplifier.

It should be borne in mind that the material presented in this section only gives an idea of the method for analysing the characte-

ristics of intrinsic noise generated by the selective circuit of an amplifier. The mechanism of noise generation depends on a number of physical and structural parameters of pure-gain (active) elements, that are not discussed here.

In conclusion, it should be pointed out that the above relations can also be used for the analysis of thermal noise produced in selective circuits. But in this case, the energy spectrum of such noise must be defined by the following formula, well-known in physics:

$$W_v(\omega) = 2kTR \quad (7.26)$$

where R is the resistance of the noise-generating resistor, $k = 1.38 \times 10^{-23}$ W s/deg is the Boltzmann constant, and T is the absolute temperature.

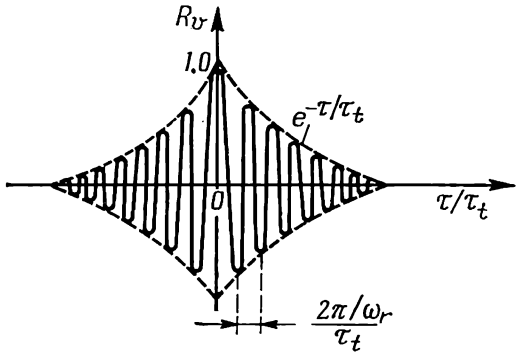


Fig. 7.7. Normalized autocorrelation function corresponding to the spectrum $W_v(\omega)$ (Fig. 7.6)

As in expression (7.12), $W_v(\omega)$ here is defined for both positive and negative frequencies. When the noise power is defined for positive frequencies only, the factor 2 must be replaced by 4.

7.3. DIFFERENTIATION OF A RANDOM FUNCTION

Let us have a stationary ergodic random process $s(t)$ with an energy spectrum $W_s(\omega)$ and autocorrelation function $B_s(\tau)$, and let it be necessary to find the characteristics of the derivative of $s(t)$. Without discussing all the conditions for the differentiability of a random function, let us restrict ourselves to the basic requirement: the energy spectrum $W_s(\omega)$ must decrease faster than $1/\omega^2$ as $\omega \to \infty$, so that

$$\int_{-\infty}^{\infty} \omega^2 W_s(\omega)\, d\omega < \infty \quad (7.27)$$

This condition is satisfied in solving most practical problems, because the energy spectrum $W_s(\omega)$ is formed by a physical circuit whose transfer function decreases faster than $1/\omega$ (and the square of its magnitude decreases faster than $1/\omega^2$) as $\omega \to \infty$. White noise with an infinitely wide spectrum does not obey condition (7.27) but one usually considers noise with a limited spectrum.

Considering condition (7.27) to be satisfied, let us discuss the transmission of a random signal $s(t)$ through an ideal differentiation circuit whose transfer function $\mathbf{K}(i\omega) = i\omega\tau_0$ [see (6.17)].

Using expressions (7.1) and (7.2), we may write

$$W_{out}(\omega) = K^2(\omega)\, W_s(\omega) = \tau_0^2 \omega^2 W_s(\omega) \tag{7.28}$$

$$B_{out}(\tau) = \tau_0^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 W_s(\omega)\, e^{i\omega\tau}\, d\omega \tag{7.29}$$

The variance of the process at the output of the system

$$\sigma_{out}^2 = \frac{\tau_0^2}{2\pi} \int_{-\infty}^{\infty} \omega^2 W_s(\omega)\, d\omega \tag{7.30}$$

Let us consider the following example. Let the spectrum of the process at the input of a differentiation device be uniform in the frequency band $-f_1 \leqslant f \leqslant f_1$:

$$W_s(\omega) = \begin{cases} W_0 & \text{for } |\omega| \leqslant 2\pi f_1 = \Delta\omega_1 \\ 0 & \text{for } |\omega| > 2\pi f_1 = \Delta\omega_1 \end{cases}$$

The autocorrelation function of such a process [see (4.41)] is

$$B_s(\tau) = W_0 2f_1 (\sin \Delta\omega_1\tau)/\Delta\omega_1\tau$$

and the variance,

$$\sigma_s^2 = B_s(0) = W_0 2f_1 \tag{7.31}$$

The normalized autocorrelation function is

$$R_s(\tau) = (\sin \Delta\omega_1\tau)/\Delta\omega_1\tau \tag{7.32}$$

After differentiating, we obtain, respectively,

$$W_{out}(\omega) = \tau_0^2 \omega^2 W_0 = (\Delta\omega_1\tau_0)^2 (\omega/\Delta\omega_1)^2 W_0$$

$$B_{out}(\tau) = \tau_0^2 W_0 \frac{1}{\pi} \int_0^{\Delta\omega_1} \omega^2 \cos\omega\tau\, d\omega$$

$$= (\Delta\omega_1\tau_0)^2 \sigma_s^2 \frac{2\Delta\omega_1\tau \cos\Delta\omega_1\tau + (\Delta\omega_1^2\tau^2 - 2)\sin\Delta\omega_1\tau}{(\Delta\omega_1\tau)^3}$$

$$\sigma_{out}^2 = B_{out}(0) = \frac{1}{3}(\Delta\omega_1\tau_0)^2 \sigma_s^2 \tag{7.33}$$

$$R_{out}(\tau) = 3y(\tau) \tag{7.34}$$

where $y(\tau)$ is the fraction in the expression for $B_{out}(\tau)$.

The graphs of the functions $W_s(\omega)$ and $W_{out}(\omega)$, and also of the functions $R_s(\tau)$ and $R_{out}(\tau)$ are shown in Fig. 7.8*a* and *b*; the parameter $\Delta\omega_1\tau_0$ is taken equal to unity. From these graphs it is clear that the differentiation causes attenuation of the lower frequencies of the original process. A relative increase in the higher frequencies results in a more pronounced oscillation of the autocorrelation function (Fig. 7.8*b*).

Now, let us consider the transmission of the same random signal through a real differentiation *R-C* circuit (Fig. 6.5*b*). The square of

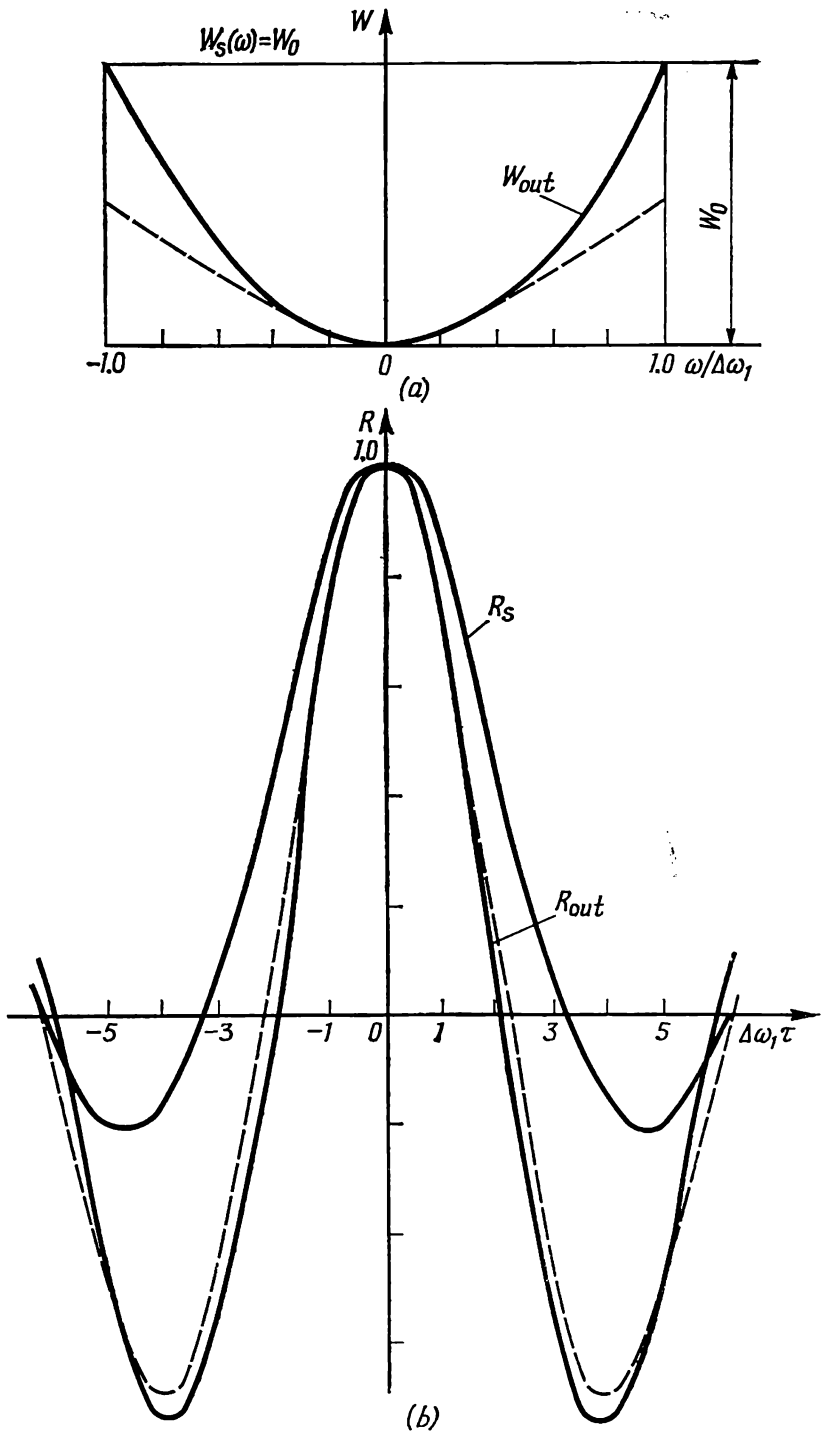


Fig. 7.8. (*a*) Energy spectra and (*b*) autocorrelation functions at the input and output of a differentiating circuit (continuous line: at the output of an ideal circuit; dashed line: at the output of an *R-C* circuit)

the transfer function of the differentiation circuit, in accordance with (6.19), is

$$K^2(\omega) = \omega^2\tau_0^2/(1 + \omega^2\tau_0^2), \quad \tau_0 = RC$$

Thus, the energy spectrum at the circuit output

$$W_{out}(\omega) = \frac{\omega^2\tau_0^2}{1+\omega^2\tau_0^2} W_0 = \frac{(\Delta\omega_1\tau_0)^2 (\omega/\Delta\omega_1)^2}{1+(\Delta\omega_1\tau_0)^2 (\omega/\Delta\omega_1)^2} W_0 \qquad (7.35)$$

The graph of $W_{out}^2(\omega)$ for $\Delta\omega_1\tau_0 = 1$ is shown in Fig. 7.8*a* by a dashed line.

The autocorrelation function

$$B_{out}(\tau) = \tau_0^2 W_0 \frac{1}{\pi} \int_0^{\Delta\omega_1} \frac{\omega^2 \cos\omega\tau}{1+\omega^2\tau_0^2} d\omega$$

$$\sigma_{out}^2 = B_{out}(0) = \tau_0^2 W_0 \frac{1}{\pi} \int_0^{\Delta\omega_1} \frac{\omega^2}{1+\omega^2\tau_0^2} d\omega$$

$$= \frac{W_0}{\tau_0} \frac{1}{\pi} [\Delta\omega_1\tau_0 - \arctan(\Delta\omega_1\tau_0)] \qquad (7.36)$$

The result of calculation of the normalized autocorrelation function $R_{out}(\tau) = B_{out}(\tau)/\sigma_{out}$ is shown in Fig. 7.8*b* by a dashed line (for $\Delta\omega_1\tau_0 = 1$).

We may consider that for $\Delta\omega_1\tau_0 \leqslant 1$ the physical *R-C* circuit differentiates the random process in question fairly accurately.

7.4. INTEGRATION OF A RANDOM FUNCTION

To reveal some specific features of the integration of a random function, let us first discuss the transmission of a stationary random process through a physical integration *R-C* circuit (Fig. 6.6*b*).

Let the input of this circuit be acted upon, starting at the moment $t = -\infty$, by a random function $s(t)$ with an energy spectrum $W_s(\omega)$ and autocorrelation function $B_s(\omega)$. Considering the process at the output to be steady-state, we can find $W_{out}(\omega)$ and $B_{out}(\tau)$ by means of expressions (7.1) and (7.2), substituting therein

$$K^2(\omega) = 1/[1 + (\omega\tau_0)^2]$$

[see (6.20)].

Thus,

$$W_{out}(\omega) = K^2(\omega) W_s(\omega) = W_s(\omega)/(1 + \omega^2\tau_0^2) \qquad (7.37)$$

$$B_{out}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{W_s(\omega) \cos\omega\tau}{1+\omega^2\tau_0^2} d\omega \qquad (7.38)$$

Let us consider two particular cases: $\overline{s(t)} = 0$ and $\overline{s(t)} \neq 0$. In the first case, the energy spectrum $W_s(\omega)$ includes no term with a

δ-function [see (4.35) through (4.37)]; setting $W_s(\omega) = W_0 = \text{const}$ (white noise), we obtain the autocorrelation function

$$B_{out}(\tau) = \frac{W_0}{\tau_0^2}\frac{1}{\pi}\int_0^\infty \frac{\cos\omega\tau}{1/\tau_0^2+\omega^2}\,d\omega = \frac{W_0}{2\tau_0}e^{-|\tau|/\tau_0} \qquad (7.39)$$

and the variance

$$\sigma_{out}^2 = W_0/2\tau_0 = W_0/2RC \qquad (7.40)$$

In the second case ($\overline{s(t)} \neq 0$), where, in accordance with (4.35), the energy spectrum

$$W_s(\omega) = [\overline{s(t)}]^2\,2\pi\delta(\omega) + W_\sim(\omega)$$

with $W_\sim(\omega) = W_0 = \text{const}$ (as in the preceding case), the autocorrelation function and variance are given by

$$B_{out}(\tau) = [\overline{s(t)}]^2\,2\pi\frac{1}{2\pi}\int_{-\infty}^{\infty}\delta(\omega)\frac{\cos\omega\tau}{1+\tau_0^2\omega^2}\,d\omega$$

$$+\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{W_\sim(\omega)\cos\omega\tau}{1+\tau_0^2\omega^2}\,d\omega = [\overline{s(t)}]^2 + \frac{W_0}{2\tau_0}e^{-|\tau|/\tau_0} \qquad (7.41)$$

$$\sigma_{out}^2 = W_0/2\tau_0 = W_0/2RC \qquad (7.42)$$

From these relations it is clear that under steady-state conditions the process at the output of a physical integration circuit is as stationary as that at the input.

It is different with exact mathematical integration to which corresponds the unrealizable transfer function

$$K(i\omega) = 1/i\omega\tau_0 \text{ [see (6.18)]}$$

In this case, the condition for the integrability of a random process assumes the following form:

$$\int_{-\infty}^{\infty}\frac{W_s(\omega)}{\omega^2}\,d\omega < \infty \qquad (7.43)$$

While the condition (7.27) for the differentiability of a random function imposes the requirement of a sufficiently fast decrease in $W_s(\omega)$ as $\omega \to \infty$, in the case of integration, a similar requirement applies to the behaviour of $W_s(\omega)$ as $\omega \to 0$.

The integration of a stationary process $s(t)$ with $W_s(0) \neq 0$ yields a *nonstationary* process with an infinitely increasing variance.

If $\overline{s(t)} \neq 0$, the mathematical expectation of the process at the output also increases infinitely.

It should be borne in mind that an ideal integrator may be regarded as a filter with an infinitely narrow passband. In such a filter,

it takes an infinitely long time to reach steady state. Therefore, the statistical characteristics of the integral of a random process depend essentially on the limits of the integral, i.e., on the integration time.

7.5. NORMALIZATION OF RANDOM PROCESSES IN NARROW-BAND LINEAR CIRCUITS

Let a stationary random process with a distribution other than normal act at the input of a linear circuit with constant parameters. If the correlation interval of this process is less than the time constant of the circuit (i.e., if the width of the energy spectrum is greater than the passband of the circuit), the distribution of the random process at the output approaches the normal one. The narrower the passband of the circuit, the stronger the normalization effect. Let this statement be illustrated by two examples.

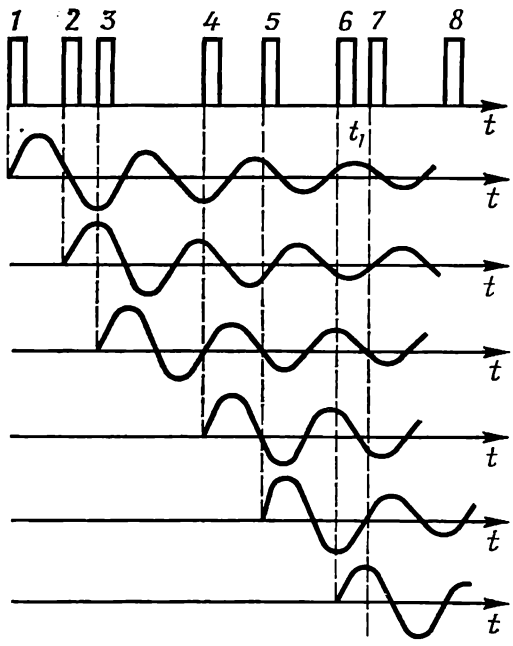


Fig. 7.9. Responses of a tank circuit to the individual pulses of a random sequence

First, we shall consider the effect of a train of short nonoverlapping pulses, randomly distributed along the time axis, on a high-Q tank circuit (Fig. 7.9), the time constant τ_t of the circuit being large compared to the mean value of the intervals between the pulses. The voltage across the circuit at some instant t_1 is a sum of free oscillations caused by the preceding pulses which had not enough time to decay completely by the given instant. The narrower the passband of the circuit, the longer the time the free oscillations persist and, consequently, the larger the number of commensurable but uncorrelated terms which take part in the formation of resultant voltage at the moment t_1. In accordance with the central limit theorem, these premises are sufficient for the distribution law to be considered approximately normal.

With the spectral approach, the normalization effect can be explained as follows. The oscillation spectrum in the tank circuit is a sum of the spectra of the individual pulses of the input train. The phases of the spectral components within each of these partial spectra are

completely correlated, while the phases of the components of different spectra are not correlated at all (due to the random distribution of the pulses along the time axis). The narrower the passband of the circuit, the less important the phase correlation in the partial spectra.

Now, let us give another example of the normalization phenomenon in a narrow-band circuit. Let this circuit be excited by a continuous constant-amplitude oscillation modulated in frequency accor-

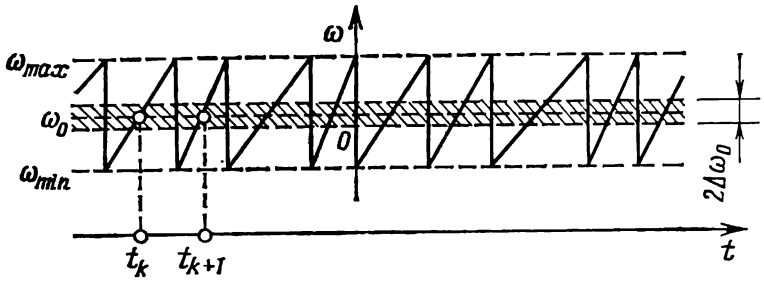


Fig. 7.10. Oscillation frequency variation according to a saw-tooth law with a random period

ding to a sawtooth law with a random period (Fig. 7.10). Each time the frequency passes through the passband $2\Delta\omega_0$ of the circuit, there arises a free oscillation whose amplitude is inversely proportional to the slope of the "saw". Since the moments of intersection of the passband are distributed at random along the time axis, the pulse train formed by free oscillations also has random intervals (t_k, t_{k+1}).

With a slow frequency sweep, when the intervals are large compared to the time constant τ_t of the tank circuit, the free oscillations do not overlap. Suppose that τ_t is large compared to the average interval T_{av}. Then, at any instant of time, a lot of oscillations with random and mutually independent phases and amplitudes are superimposed. In this case, the input oscillation, whose distribution is defined by formula (4.25), transforms into a random function with a distribution close to normal (an instantaneous frequency variation has no effect on the one-dimensional distribution law of an r-f oscillation with a constant amplitude). The greater the τ_t compared to T_{av}, the more complete the normalization.

Taking into account that for a solitary tank circuit $\Delta\omega_0\tau_t = 1$ and the average frequency of the "saw" is $F_{av} = 1/T_{av}$, the condition for normalization may be written in the form of the following inequality:

$$F_{av} \gg \Delta\omega_0 \tag{7.44}$$

In wide-band linear circuits, under certain conditions there may occur an effect that is the opposite of the above effect of normalization: the distribution of the process at the output of the circuit may

differ from normal to a greater extent than at the input. Let us give a simple example of such an effect.

Let a set of comparatively long pulses be applied to the input of a differentiator, the distribution of these pulses being close to normal. After differentiation, each input pulse is transformed into a pair of very short pulses at the output of the circuit, these pulses corresponding to the edges of the pulse. The number of the overlapping pulses at the output is thus reduced, and the approximation of the normal law at the output becomes worse than at the input. Such an effect is sometimes called the "denormalization" of the process.

It should be pointed out that the above effect does not contradict the fact that in any linear circuit a normal process retains its normal distribution law. If in the above example the average number of pulses per unit time is brought to infinity (which is required for obtaining strictly normal distribution), the process at the circuit output will also be normal with any compression of the pulses that can be effected in any physically realizable circuit.

7.6. DISTRIBUTION OF A SUM OF HARMONIC OSCILLATIONS WITH RANDOM PHASES

This problem is gaining in importance in connection with the wide use of spectral representation of signals and random processes. From the central limit theorem (see the footnote on page 277) it follows that the sum of a sufficiently large number of harmonic oscillations with random and mutually independent amplitudes and phases forms a random function obeying the normal distribution law.

In practice, one often encounters problems involving a comparatively small number of terms, so that the conditions of the central limit theorem are not satisfied and complete normalization is not obtained.

To find the dependence of the rate of normalization on the number of random terms (not obeying the normal distribution law) we may use the *method of characteristic functions.*

In the probability theory, the characteristic function $\theta(\eta)$ of a random variable x, or the characteristic function of the given distribution $p(x)$ is the mean value of the function $e^{i\eta x}$ i.e.,

$$\theta(\eta) = \langle e^{i\eta x} \rangle \tag{7.45}$$

where η is a real variable.

With the probability density $p(x)$ being specified, the mean value of $e^{i\eta x}$ can be found from the expression

$$\theta(\eta) = \langle e^{i\eta x} \rangle = \int_{-\infty}^{\infty} e^{i\eta x} p(x)\, dx \tag{7.46}$$

The right-hand side of this equation is the Fourier transform of the function $p(x)$. Therefore, if we know the characteristic function of any random variable x, the probability density $p(x)$ can be found by means of the inverse [with respect to (7.46)] Fourier transform

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \theta(\eta)\, e^{-i\eta x}\, d\eta \tag{7.47}$$

The characteristic function has some important properties that allow one to determine the parameters of the distribution $p(x)$. It is especially important that in the case of N mutually independent terms $x_1, x_2, \ldots, x_N$, the characteristic function of the sum of these terms has the following form

$$\theta(\eta) = \langle e^{i\eta x} \rangle = \langle e^{i\eta(x_1+x_2+\ldots+x_n)} \rangle = \langle e^{i\eta x_1} \rangle \langle e^{i\eta x_2} \rangle \ldots \langle e^{i\eta x_N} \rangle$$
$$= \theta_1(\eta)\, \theta_2(\eta) \ldots \theta_N(\eta) \tag{7.48}$$

i.e., the characteristic function of a sum of independent random variables is equal to the product of the characteristic functions of these variables.

In particular case where all the terms have the same distribution and hence, the same characteristic function, we have

$$\theta_N(\eta) = [\theta_1(\eta)]^N \tag{7.49}$$

For the most important and the most widely spread normal distribution law

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right)$$

and the characteristic function, in accordance with (7.46), is

$$\theta(\eta) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \exp(i\eta x)\, dx$$

Using transformations identical to (2.75) and (2.77), we get*

$$\theta(\eta) = \exp\left(-\frac{\sigma^2\eta^2}{2}\right) \tag{7.50}$$

Thus, with the normal distribution, the graph of the characteristic function with respect to η has the same shape as the graph of

* In the general case, where the mean of a random variable is nonzero and

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \exp\left[-\frac{(x-\bar{x})^2}{2\sigma_x^2}\right]$$

the characteristic function is given by

$$\theta(\eta) = \exp(i\bar{x}\eta - \sigma_x^2\eta^2/2)$$

(see, for example, [7]).

the probability density with respect to x. Therefore, the degree of similarity between the characteristic function of a random variable and the function defined by expression (7.50) defines the degree of similarity between the distribution of this variable and the normal distribution.

Let us use expressions (7.46) through (7.50) to find the distribution of a sum of several sinusoidal waves with random phases.

For a sample x taken from a harmonic oscillation with an amplitude A_0 and a random phase, the probability density [see (4.25)] is

$$p(x) = 1/\pi \sqrt{A_0^2 - x^2}, \qquad -A_0 \leqslant x \leqslant A_0 \tag{7.51}$$

From expression (7.46) we find the characteristic function of this distribution

$$\theta_1(\eta) = \frac{1}{\pi} \int_{-A_0}^{A_0} \frac{e^{i\eta x}}{\sqrt{A_0^2 - x^2}}\, dx \tag{7.52}$$

Substituting $e^{i\eta x} = \cos \eta x + i \sin \eta x$ and taking into account that $\sin \eta x / \sqrt{A_0^2 - x^2}$ is an odd function of x, we obtain

$$\theta_1(\eta) = \frac{2}{\pi} \int_0^{A_0} \frac{\cos \eta x}{\sqrt{A_0^2 - x^2}}\, dx = J_0(A_0\eta) \tag{7.53}$$

where J_0 is the Bessel function of the first kind and of zero order.

For the sample taken from a sum of N harmonic oscillations with equal amplitudes A_0/N but with random and mutually independent phases, the characteristic function, in accordance with (7.49), is

$$\theta_N(\eta) = [J_0(A_0\eta/\sqrt{N})]^N \tag{7.54}$$

The amplitude of each sinusoid is taken equal to $A_0/\sqrt{N}$ so that the variance of the sum, which is equal to $(N/2)(A_0/\sqrt{N})^2$, can remain constant as the number of the sinusoids is increased.

Figure 7.11 shows the characteristic functions for various values of N. When $N \geqslant 4$ or 5, the function $\theta_N(\eta)$ quickly approaches the limiting curve $N \to \infty$, which corresponds to the normal distribution of the sum.

To find the probability density of the sum of N harmonic oscillations, in accordance with expression (7.47), it is necessary to calculate the integral

$$p_N(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \theta_N(\eta)\, e^{-i\eta x}\, d\eta$$

$$= \frac{1}{\pi} \int_0^{\infty} \left[J_0\left(\frac{A_0\eta}{\sqrt{N}}\right) \right]^N \cos \eta x\, d\eta \tag{7.55}$$

When $N = 1$, we have the initial expression for $p(x)$ for one sinusoid [formula (7.51)], while when $N = 3 \ldots 4$, the functions $p_N(x)$ have the form shown in Fig. 7.12, which is plotted for the case $A_0 = 1$ (the integrals were calculated approximately). The continuous line shows the function $p_N(x)$ for the normal distribution ($N \to \infty$).

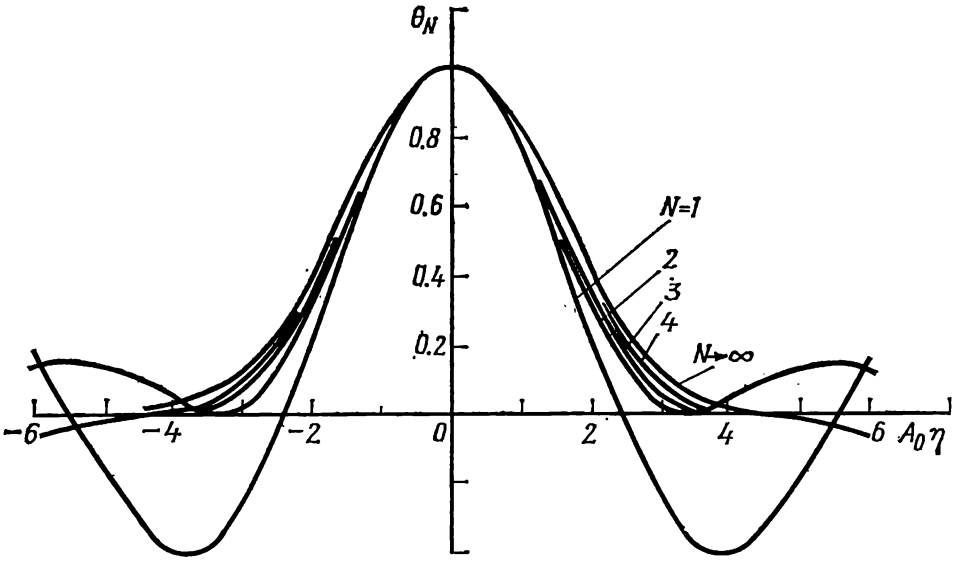


Fig. 7.11. Characteristic functions for the sum of N harmonic oscillations with random phases

The results obtained show that the summation of even five or six harmonic oscillations with random and mutually independent phases yields a stationary random process close to normal.

This holds for the values of x within the region $|x| < \sqrt{N}$ (when $A_0 = 1$). For higher values of $|x|$, the probability density $p_N(x) = 0$, whereas for normal distribution $p(x) \neq 0$. Thus, for a finite number N of terms, a discrepancy between $p_N(x)$ and $p_\infty(x)$ is unavoidable in the "tails" of the distribution curve.

The above-mentioned stationarity of the sum stems from the stationarity of the original process (a set of harmonic oscillations with random phases). Then, since the probability density $p(x)$ defined by formula (7.51) is independent of frequency, $p_N(x)$ is also independent of frequency.

However, it should be borne in mind that the summation of harmonic oscillation of the same frequency also yields a stationary but nonergodic process. In this case, each realization of the resultant process is a harmonic oscillation differing from the other realizations only in amplitude and phase (depending on the phases with

which the original N oscillations have added in the given realization). When averaging x^2 "along the process", each kth realization has its own value of $\overline{x_k^2}$, which does not coincide with the true variance σ_x^2 obtained by ensemble averaging (i.e., "across the process").

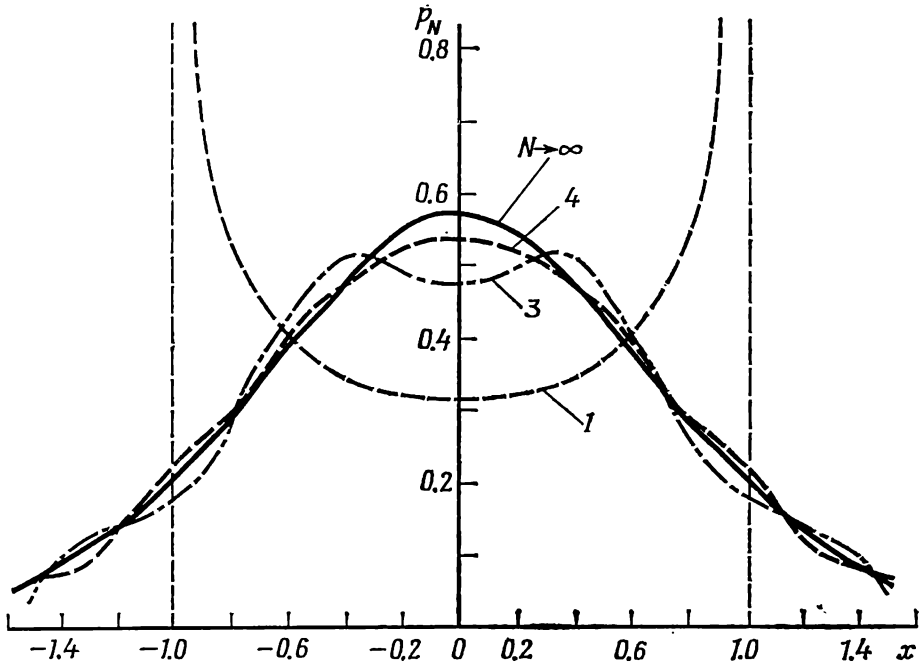


Fig. 7.12. Probability density of the sum of N harmonic oscillations with random phases (Fig. 7.11)

The summation of harmonic oscillations of random epoch angles and different frequencies yields not only stationary but ergodic process (for sufficiently large values of N).

Thus, the summation of a sufficiently large number of uncorrelated harmonic oscillations yields a stationary process close to normal.

Chapter 8

NONLINEAR CIRCUITS AND METHODS OF THEIR ANALYSIS

8.1. NONLINEAR ELEMENTS

Principal conversions in radio circuits are effected either by means of nonlinear circuits or by means of linear circuits with variable parameters. However, linear circuits with variable parameters are realized with the aid of nonlinear elements (for example, the capacitance of a *p-n* junction in a semiconductor diode), while some parametric circuits operate under essentially nonlinear conditions (e.g., a parametric oscillator). We may therefore say that the theory of the majority of actual radio devices is based on the properties of nonlinear elements and circuits. In this chapter we present some examples of nonlinear elements.

One should distinguish between *resistive* nonlinear elements (resistors) and *reactive* nonlinear elements (inductors, capacitors).

Radio circuits and devices are featured by a wide use of resistive nonlinear elements, such as electron tubes, semiconductor and other

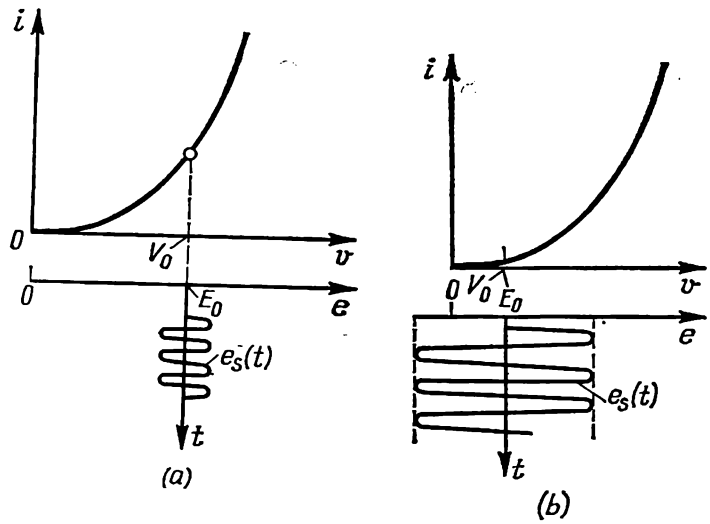


Fig. 8.1. (*a*) Linear and (*b*) nonlinear operation of a device with a nonlinear current-voltage characteristic

devices for amplifying and converting signals, which have a nonlinear current-voltage characteristic. A most important parameter of a resistive nonlinear element is the slope of its characteristic curve.

The following two definitions of the slope of the characteristic curve should be distinguished: (a) the slope at a given operating point in the case of a weak signal (differential slope) and (b) the slope in the case of a strong harmonic oscillation (average slope).

The first definition of the slope, corresponding to linear operating conditions of a device (Fig. 8.1*a*), was discussed in Ch. 5, where this slope was defined by expression [see (5.37′) and (5.40)] of the form

$$S = a_1 = \left(\frac{di}{dv}\right)_{v=V_0} \tag{8.1}$$

where V_0 was taken equal to V_{BE0} (for a transistor).

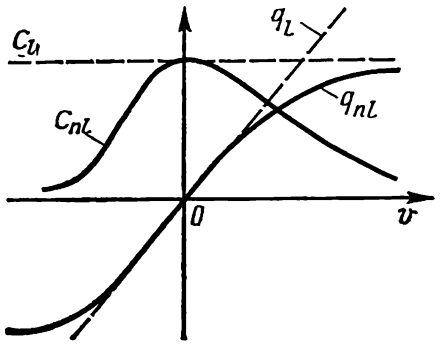


Fig. 8.2. Volt-coulomb and volt-farad characteristic of a linear (continuous line) and nonlinear (dashed line) capacitor

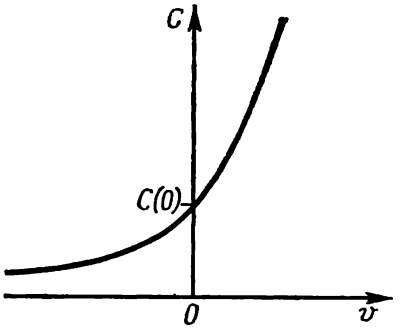


Fig. 8.3. Approximate view of the volt-farad characteristic of a semiconductor diode

The second definition of the slope corresponds to essentially nonlinear operating conditions of a device (Fig. 8.1*b*) and can be given only with due regard for the shape of the current-voltage characteristic of the nonlinear element within the range of change of the stimulus (it will be done in Sec. 8.4).

Any device having a nonlinear volt-coulomb characteristic $q(v)$ may serve as an example of a nonlinear circuit.

Figure 8.2 shows the volt-coulomb characteristic $q_l(v)$ and the volt-farad characteristic $C_l = q_l(v)/v = \text{const}$ for an ordinary linear capacitor, and similar characteristics q_{nl} and $C_{nl} = q_{nl}(v)/v$ for a nonlinear capacitor.

No matter what the dependence of C_{nl} on the charge q_{nl}, as in the case of a linear capacitor, we have

$$q_{nl}(v) = C_{nl}(v)\, v \tag{8.2}$$

In the text below, a nonlinear capacitor will be denoted $C(v)$.

If the voltage applied to a capacitor $C(v)$ changes in time, the current through the capacitor can be found by means of either of

the two equivalent expressions:

$$i(t) = \frac{dq(v)}{dt} = \frac{dq(v)}{dv}\frac{dv}{dt} \tag{8.3}$$

$$i(t) = \frac{d[C(v)\,v]}{dt} = v\frac{dC(v)}{dt} + C(v)\frac{dv}{dt} = \left[v\frac{dC(v)}{dv} + C(v)\right]\frac{dv}{dt}$$

If the voltage v changes within narrow limits in the vicinity of the point $v = V_0$, the capacitance may be represented as

$$C_0 \approx \left.\frac{dq(v)}{dv}\right|_{v=V_0} \tag{8.4}$$

The capacitance thus found is sometimes called *differential.*

Figure 8.3 shows an approximate shape of the curve $C(v)$ for a semiconductor diode.

Finally, a coil with a ferromagnetic core which is magnetically saturated by a strong current flowing through the coil is an example of a nonlinear inductance $L(i)$.

The relationship between the current i and the voltage v_L across the inductance follows from the initial expression for magnetic linkage:

$$\Phi(i) = L(i)\,i \tag{8.5}$$

It is clear that

$$v_L(t) = \frac{d\Phi(i)}{dt} = \frac{dL(i)}{dt}\frac{di}{dt}\,i + L(i)\frac{di}{dt} = \left[i\frac{dL(i)}{di} + L(i)\right]\frac{di}{dt} \tag{8.6}$$

If the voltage $v_L(t)$ across the inductance is specified, then, obviously,

$$\int v_L(t)\,dt = \Phi(t) = L(i)\,i(t),$$

and, as in the case of a linear inductance

$$i(t) = \frac{1}{L[i(t)]}\int v_L(t)\,dt$$

By the *differential inductance* is meant the quantity

$$L_0 \approx \left.\frac{d\Phi(i)}{di}\right|_{i=I_0} \tag{8.7}$$

The concepts of differential capacitance, inductance and resistance are widely used in studying the effect of comparatively weak signals on nonlinear elements. In this case, the nonlinearity of an element manifests itself only in that the quantities R_0, C_0, and L_0 depend on the value of the control voltage (or current) determining the position of the operating point on the nonlinear characteristic curve. But with respect to a *weak signal*, such an element is a *linear device with a variable parameter* (if the control voltage changes in time). The properties of such devices are discussed in Ch. 10.

8.2. APPROXIMATION OF NONLINEAR CHARACTERISTICS

For the analysis and calculation of nonlinear circuits, it is necessary to specify the current-voltage or similar characteristics of nonlinear elements in an analytic form.

Actual characteristic curves usually have a complex shape that hinders their exact description by a simple analytic expression.

Widely used in radio engineering are the methods of representing characteristics by comparatively simple functions that only approximately describe the actual characteristics. The replacement of an actual characteristic by a function that describes it approximately is called the *approximation* of characteristic.

The choice of an optimum approximation depends on the shape of the given nonlinear characteristic and the operating conditions of the nonlinear element. One of the well-known methods is the polynomial approximation.

Let us write the approximating power polynomial in the form

$$i(v) = i(V_0) + a_1(v - V_0) + a_2(v - V_0)^2 + a_3(v - V_0)^3 + \ldots \quad (8.8)$$

If the nonlinear element is a transistor, then i is the collector current and v is a voltage, e.g., across the base and emitter. For a vacuum triode or pentode, v is the voltage across the grid and anode, and i is the anode current, etc.

The coefficients a_1, a_2, a_3 are defined by the expressions

$$a_1 = \left(\frac{di}{dv}\right)_{v=V_0}, \quad a_2 = \frac{1}{2!}\left(\frac{d^2i}{dv^2}\right)_{v=V_0}, \quad a_3 = \frac{1}{3!}\left(\frac{d^3i}{dv^3}\right)_{v=V_0} \quad (8.9)$$

It is clear that a_1 is the slope of the characteristic curve at the point $v = V_0$, a_2 is the first derivative of the slope (with a factor 1/2!), a_3 is the second derivative of the slope (with a factor 1/3!), etc.

The shape of the current-voltage characteristic being specified, the magnitudes of the coefficients a_1, a_2, a_3, etc. depend essentially on the position of the operating point on the characteristic curve.

Let us consider a few typical cases that are of practical importance.

1. The operating point is situated in the initial section of the characteristic curve having the shape of a quadratic parabola (Fig. 8.4). It is assumed that the signal voltage e_s applied to the nonlinear element and superimposed on the d-c voltage $E_0 = V_0$ does not go beyond the point V_1, i.e., the origin of the characteristic.

Expression (8.8) can in this case be written as the second-degree polynomial

$$i(V_0 + e_s) = i(V_0) + a_1 e_s + a_2 e_s^2 \quad (8.10)$$

The coefficient a_1 found from expression (8.9) is often denoted as S (the slope of the characteristic curve).

The coefficient a_2 is determined from the condition that for $e_s = V_1 - V_0$, the current $i = 0$, whence

$$i(V_0) + S(V_1 - V_0) + a_2(V_1 - V_0)^2 = 0$$

Thus,

$$a_2 = -[i(V_0) + S(V_1 - V_0)]/(V_1 - V_0)^2 \qquad (8.11)$$

2. The operating point is the inflection point of the characteristic curve shown in Fig. 8.5.

At the inflection point of the curve $i = f(v)$, all the even-order derivatives are zero. Therefore, the coefficients multiplying the

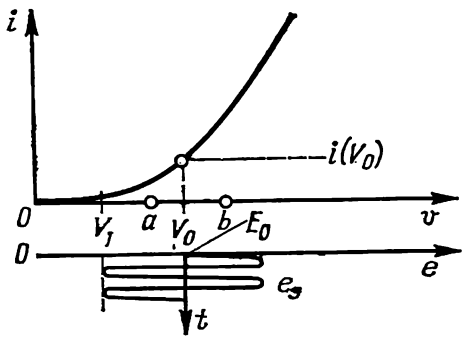


Fig. 8.4. Position of the operating point and the limits (a, b) of utilization of the current-voltage characteristic within which a second-degree polynomial approximation can be applied

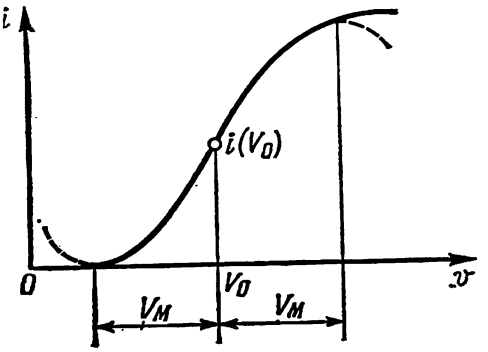


Fig. 8.5. Example of a characteristic requiring a third-degree polynomial approximation

terms with even exponents in expression (8.8) vanish, and the expression may be written in the form

$$i(v) = i(V_0) + a_1(v - V_0) + a_3(v - V_0)^3 + a_5(v - V_0)^5 + \ldots \qquad (8.12)$$

In order to simplify the analysis, one often restricts oneself to the third-degree polynomial without the quadratic term (incomplete third-degree polynomial). Replacing, as in Para. 1, $v - V_0$ by the signal voltage e_s, we obtain

$$i(V_0 + e_s) = i(V_0) + a_1 e_s + a_3 e_s^3 \qquad (8.13)$$

The respective approximated characteristic is shown in Fig. 8.5 by a dashed line. The voltage V_M corresponding to the extreme points of the approximating function and read from $v = V_0$ is sometimes called saturation voltage. The coefficient a_3 in expression (8.13) is unambiguously defined by this specific voltage and by a_1 (the slope S at the point V_0).

In fact, at the point $V_0 + V_M$, i.e., where the amplitude of the input signal is equal to V_M, the identity

$$\left(\frac{di}{de_s}\right)_{e_s=V_M} = a_1 + 3a_3V_M^2 = 0$$

is satisfied, whence

$$a_3 = -a_1/3V_M^2 = -S/3V_M^2 \qquad (8.14)$$

It should be noted that approximation (8.13) can be used when the signal voltage does not go beyond the limits $\pm V_M$.

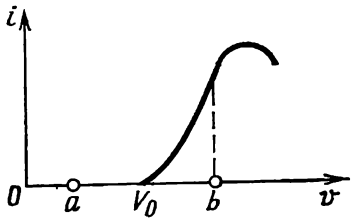


Fig. 8.6. Example of a characteristic requiring a high-degree polynomial approximation

3. The operating point is on the lower bend of the characteristic curve shown in Fig. 8.6.

If the voltage changes are so great that the section denoted by the letters *a* and *b* on the abscissa axis is used, satisfactory approximation can only be obtained by employing a fifth- or higher-degree poly-

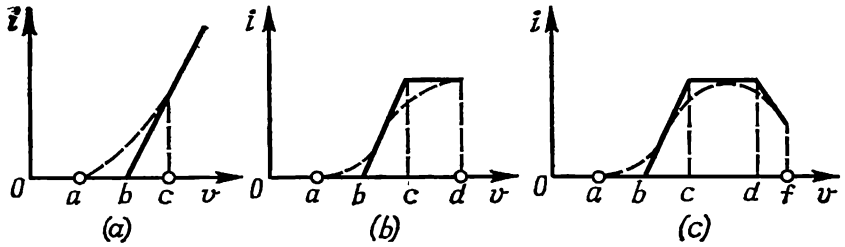


Fig. 8.7. Examples of piecewise-linear approximation of a characteristic for various limits of its utilization

nomial. In this case the analysis becomes complicated and the use of polynomial approximation for practical purposes proves ineffective.

With very high signal amplitudes, it is often turns convenient to replace the actual characteristic by an idealized broken line. Such representation of a characteristic is referred to as *piecewise-linear approximation*. Some examples of piecewise-linear approximation are given in Fig. 8.7. Figure 8.7*a* corresponds to the case where the lower bend and the linear part of the characteristic curve (section *a-c*) are used, Fig. 8.7*b* represents the case where the signal variation embra-

ces both the lower and the upper bends (section *a-d*), and Fig. 8.7*c* is the case where the signal also reaches the dropping section of the characteristic curve (section *a-f*). It should be emphasized that the *replacement of an actual linear characteristic by line-segments does not mean the linearization of the circuit.* For example, in spite of the fact that the characteristic is linear in the section *b-c*, with respect to the signal embracing the section *a-c* (Fig. 8.7*a*) the entire system is essentially nonlinear.

The piecewise-linear approximation is particularly convenient for the analysis when the lower bend of the characteristic curve is of primary importance, i.e., when one may restrict oneself to two straight lines (Fig. 8.7*a*). Where the operating section of the characteristic curve has a more complex shape, the number of the approximating sections increases and the piecewise-linear approximation looses its advantages. In such cases, approximation is sometimes effected through the use of various transcendental functions, e.g., hyperbolic tangent [4], exponential functions, and some others.

The above methods of approximation are also applicable to the corresponding characteristics of *reactive* nonlinear elements.

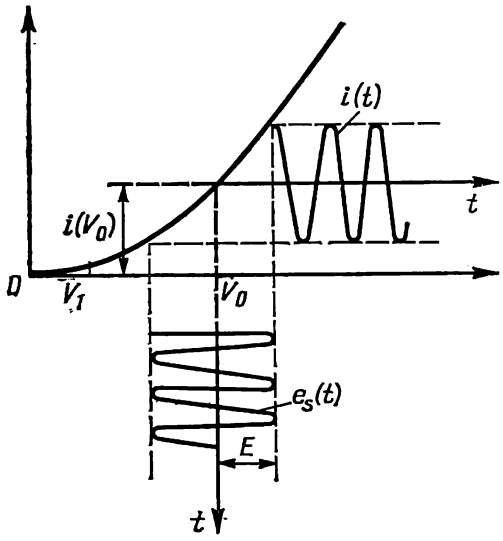


Fig. 8.8. Weakly nonlinear operation of an amplifying device

8.3. ACTION OF HARMONIC OSCILLATIONS ON CIRCUITS WITH LAG-FREE NONLINEAR ELEMENTS

The main properties of such circuits can be revealed from the analysis of the action of harmonic oscillations on *resistive* elements. Such an element may be represented by any amplifying device with a nonlinear current-voltage characteristic.

First, let us consider the operating conditions illustrated in Fig. 8.8, under which the signal voltage $e_s(t)$ does not go beyond the point V_1 and the current-voltage characteristic $i(v)$ is satisfactorily approximated by polynomial (8.8). Substituting $v = e_s(t) = E\cos\omega_1 t$ into (8.8), we obtain

$$i(t) = i(V_0) + a_1 E\cos\omega_1 t + a_2 E^2\cos^2\omega_1 t + a_3 E^3\cos^3\omega_1 t + \ldots \quad (8.15)$$

The shape of the current $i(t)$ is shown in Fig. 8.8.

Using the trigonometric relations

$$\cos^2 x = \frac{1}{2} + \frac{1}{2}\cos 2x, \quad \cos^3 x = \frac{3}{4}\cos x + \frac{1}{4}\cos 3x,$$

$$\cos^4 x = \frac{3}{8} + \frac{1}{2}\cos 2x + \frac{1}{8}\cos 4x,$$

$$\cos^5 x = \frac{5}{8}\cos x + \frac{5}{16}\cos 3x + \frac{1}{16}\cos 5x, \text{ etc.}$$

we reduce expression (8.15) to the form

$$\begin{aligned} i(t) = &\left[i(V_0) + \frac{1}{2} a_2 E^2 + \frac{3}{8} a_4 E^4 + \ldots \right] \\ &+ \left(a_1 E + \frac{3}{4} a_3 E^3 + \frac{5}{8} a_5 E^5 + \ldots \right) \cos \omega_1 t \\ &+ \left(\frac{1}{2} a_2 E^2 + \frac{1}{8} a_4 E^4 + \ldots \right) \cos 2\omega_1 t \\ &+ \left(\frac{1}{4} a_3 E^3 + \frac{5}{16} a_5 E^5 + \ldots \right) \cos 3\omega_1 t \\ &+ \left(\frac{1}{8} a_4 E^4 + \ldots \right) \cos 4\omega_1 t + \left(\frac{5}{8} a_5 E^5 + \ldots \right) \cos 5\omega_1 t \\ &+ \ldots = I_0 + I_1 \cos \omega_1 t + I_2 \cos 2\omega_1 t + I_3 \cos 3\omega_1 t + \ldots \end{aligned} \quad (8.16)$$

From this expression one can see the following manifestations of nonlinearity of the current-voltage characteristic in the case of harmonic excitation:

—the no-load current $i(V_0)$ gets an increment due to the coefficients a_2, a_4, ... multiplying the terms of even powers in polynomial (8.8):

$$I_0 = i(V_0) + \frac{1}{2} a_2 E^2 + \frac{3}{8} a_4 E^4 + \ldots \quad (8.17)$$

—the amplitude I_1 of the harmonic of the fundamental frequency ω_1 is related to the excitation amplitude E by a nonlinear relation due to the terms of odd powers in polynomial (8.8):

$$I_1 = a_1 E + \frac{3}{4} a_3 E^3 + \ldots \quad (8.18)$$

—the current $i(t)$ contains higher harmonics with frequencies $n\omega_1$ that are multiples of the excitation frequency ω_1. The harmonics with frequencies $2\omega_1$, $4\omega_1$, ... are due to the terms of even powers in polynomial (8.8), while those with frequencies $3\omega_1$, $5\omega_1$, ... are due to the terms of odd powers in this polynomial.

The current spectrum for the case where $a_1 = 2$ mA/V, $a_2 = 0.15$ mA/V^2, $a_3 = 0.03$ mA/V^3, $i(V_0) = 10$ mA, and $E = 5$ V is shown in Fig. 8.9. In this example, one may neglect all the terms of powers higher than the second in expression (8.15).

Now let us consider the operation of the same nonlinear element under essentially nonlinear conditions (Fig. 8.10*a*) obtained by shifting the operating point V_0 to the left and correspondingly increasing the amplitude of the exciting voltage E. In this case, it is advisable to use piecewise-linear approximation of the current-voltage characteristic (see Sec. 8.2, the commentary to Fig. 8.7*a*).

With harmonic excitation, the current $i(t)$ becomes pulsed (Fig. 8.10*b*). The angle θ corresponding to the change of the current from its maximum value I_m to zero is known as the current cutoff angle*. The duration of the current pulses is equal to 2θ (Fig. 8.10*b*). From Fig. 8.10*a* it follows that

$$\cos\theta = (V_1 - V_0)/E \tag{8.19}$$

The current amplitude is

$$I_m = a_1[E - (V_1 - V_0)] = a_1 E(1 - \cos\theta) \tag{8.20}$$

where a_1 is the slope of the linear part of the current-voltage characteristic [see expression (8.9)].

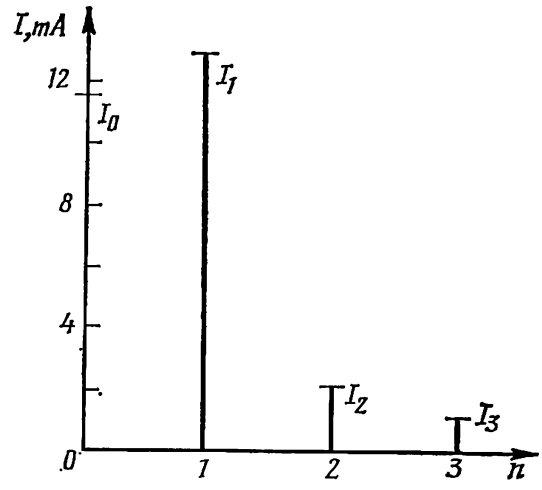


Fig. 8.9. Current spectrum for the case shown in Fig. 8.8

With harmonic excitation of the nonlinear element, the shape of the current pulse within the range $-\theta < \omega t < \theta$ is close to that of the cutoff cosine curve and, if we ignore the curvature of the current-voltage characteristic at the lower bend (Fig. 8.10*a*), the instantaneous value of the current may be given by

$$i(t) = I'_m(\cos\omega t - \cos\theta), \quad -\theta < \omega t < \theta \tag{8.21}$$

The symbol I'_m stands for the amplitude of a pulse, which would be obtained at $\theta = \pi/2$.

Since the amplitude I_m of the actual pulse corresponds to the instant $\omega t = 0$, the following relation holds:

$$I_m = i(0) = I'_m(1 - \cos\theta)$$

whence

$$I'_m = I_m/(1 - \cos\theta)$$

* The cutoff angle is equal to half the conduction (operating) angle.—*Editor.*

Substituting this expression into (8.21), we finally obtain

$$i(t) = \frac{I_m}{1-\cos\theta}(\cos\omega t - \cos\theta), \quad -\theta < \omega t < \theta \tag{8.22}$$

From this expression one can easily find the Fourier-series coefficients for the periodic pulse train shown in Fig. 8.11. Since the

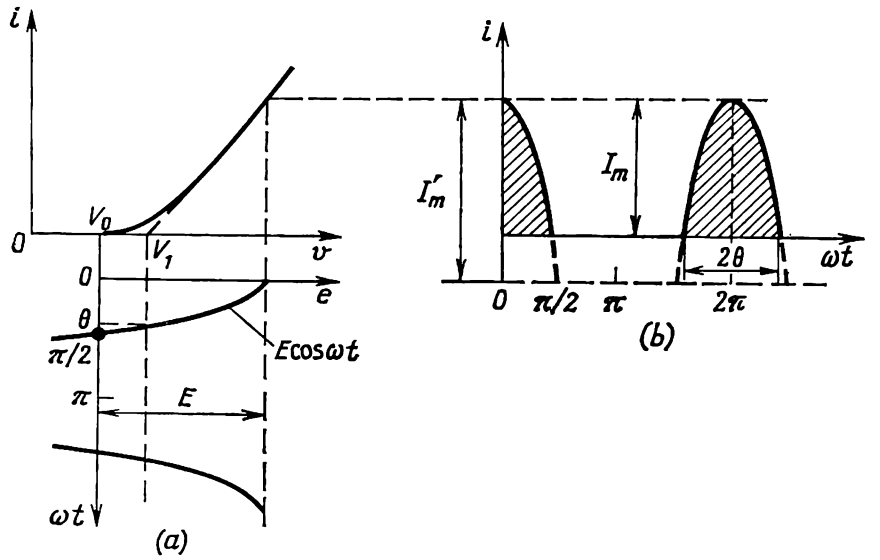


Fig. 8.10. Essentially nonlinear operation of an amplifying device

function $i(t)$ is even with respect to t [see (8.22)] the Fourier series includes cosine terms only. Using formulas (2.24) and (2.32), we find

$$I_0 = \frac{1}{2\pi}\int_{-\theta}^{\theta} i(t)\, d(\omega t)$$

$$= \frac{I_m}{\pi(1-\cos\theta)}\int_0^{\theta}(\cos\omega t - \cos\theta)\, d(\omega t) = I_m \frac{\sin\theta - \theta\cos\theta}{\pi(1-\cos\theta)} \tag{8.23}$$

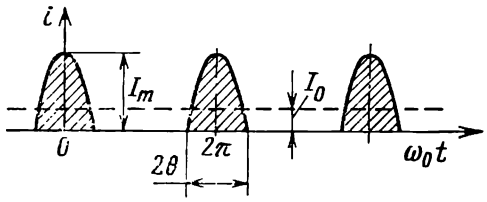


Fig. 8.11. Pulsed current corresponding to the case shown in Fig. 8.10

$$I_1 = \frac{1}{\pi}\int_{-\theta}^{\theta} i(t)\cos\omega t\, d(\omega t)$$

$$= \frac{2I_m}{\pi(1-\cos\theta)}\int_{0}^{\theta}(\cos\omega t - \cos\theta)\cos\omega t\, d(\omega t) = I_m\frac{\theta - \sin\theta\cos\theta}{\pi(1-\cos\theta)} \tag{8.24}$$

In a similar way, one can obtain the general expression for the amplitude of the nth harmonic:

$$I_n = I_m\frac{2(\sin\theta\cos\theta - n\cos n\theta\sin\theta)}{\pi n(n^2-1)(1-\cos\theta)} \tag{8.25}$$

The ratios

$$\left.\begin{aligned}
\alpha_0(\theta) &= \frac{I_0}{I_m} = \frac{\sin\theta - \theta\cos\theta}{\pi(1-\cos\theta)}\\
\alpha_1(\theta) &= \frac{I_1}{I_m} = \frac{\theta - \sin\theta\cos\theta}{\pi(1-\cos\theta)}\\
\alpha_2(\theta) &= \frac{I_2}{I_m}\\
&\cdots\cdots\cdots\cdots\\
\alpha_n(\theta) &= \frac{I_n}{I_m}
\end{aligned}\right\} \tag{8.26}$$

are the coefficients, respectively, of the d-c component, the first harmonic, etc. (Berg functions).

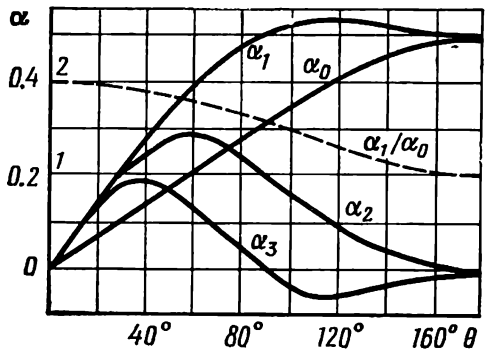


Fig. 8.12. Fourier-series coefficients for a pulsed current as a function of the cutoff angle θ

The graphs of the coefficients α_0, α_1, α_2, . . ., and also, of the ratio $\gamma_1 = \alpha_1/\alpha_0$ as a function of the cutoff angle varying from $\theta = 0$ to $\theta = 180°$ are shown in Fig. 8.12. When $\theta = 0$, there is no current at all (the nonlinear element is nonconductive during the entire period) and when $\theta = 180°$, there is no current cutoff, and the operating conditions become linear. The current spectra for several values of the cutoff angle are shown in Fig. 8.13 (for $I_m = 1$).

Examining the graphs of the functions $\alpha_n(\theta)$ and the current spectra, one can draw important conclusions. It stands out that in operation with a cutoff angle smaller than 180°, the ratio of the amplitude I_1 of the first harmonic to the d-c component I_0 is higher than unity, while under linear conditions this ratio is much lower

than unity. It is clear that with a decrease in θ the ratio

$$\gamma_1 = \frac{\alpha_1}{\alpha_0} = \frac{I_1}{I_0} = \frac{\theta - \sin\theta\cos\theta}{\sin\theta - \theta\cos\theta} \tag{8.27}$$

increases. Furthermore, with an increase in the harmonics' serial number the maxima of the functions $\alpha_n(\theta)$ shift into the region of low values of θ. All these circumstances materially influence the

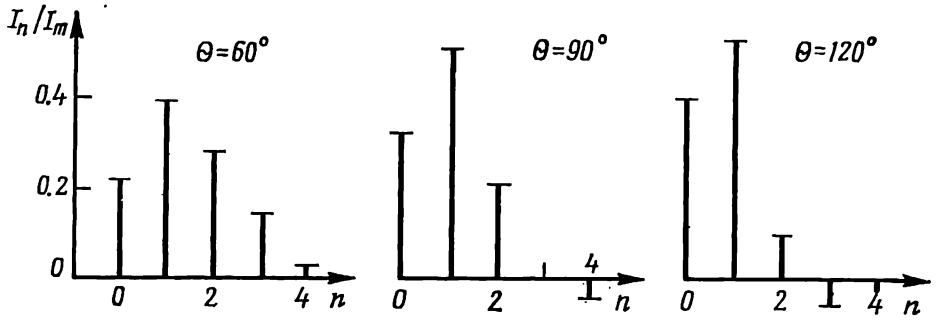


Fig. 8.13. Spectra of a pulsed current for several values of the cutoff angle θ

selection of operating conditions for nonlinear elements used for amplifying oscillations, multiplying frequency, and effecting some other conversions which are dealt with in the next sections of this chapter.

Now let us consider the effect of the *biharmonic* oscillation

$$e_s(t) = E_1 \cos\omega_1 t + E_2 \cos\omega_2 t \tag{8.28}$$

on a nonlinear resistive element.

To simplify the analysis, let us restrict ourselves to the examination of a slightly nonlinear operation (Fig. 8.8), in which case it is sufficient to take into account only the linear and quadratic terms in polynomial (8.8).

Substituting (8.28) into series (8.8) yields the following results:

—for the linear term of the series

$$a_1 e_s(t) = a_1 E_1 \cos\omega_1 t + a_2 E_2 \cos\omega_2 t \tag{8.29}$$

—for the quadratic term of the series

$$a_2 e_s^2(t) = a_2 (E_1 \cos\omega_1 t + E_2 \cos\omega_2 t)^2 = a_2 E_1^2 \cos^2\omega_1 t + a_2 E_2^2 \cos^2\omega_2 t$$
$$+ 2a_2 E_1 E_2 \cos\omega_1 t \cos\omega_2 t = \frac{1}{2} a_2 (E_1^2 + E_2^2) + \frac{1}{2} a_2 E_1^2 \cos 2\omega_1 t$$
$$+ \frac{1}{2} a_2 E_2^2 \cos 2\omega_2 t + a_2 E_1 E_2 [\cos(\omega_1 + \omega_2) t + \cos(\omega_1 - \omega_2) t] \tag{8.30}$$

The first term, that is independent of time, defines the increment of the *d-c current*. The terms of frequencies $2\omega_1$ and $2\omega_2$ are the *se-*

cond harmonics of the corresponding components of the input signal, and the terms of frequencies $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$ are the *combination* oscillations.

In a more general case, after making similar transformations with the cubic term $a_3 e_s^3(t)$, we find that this term introduces into the spectrum the following: ω_1, ω_2—fundamental frequencies; $3\omega_1$, $3\omega_2$—third harmonics; $\omega_1 + 2\omega_2$, $|\omega_1 - 2\omega_2|$, $2\omega_1 + \omega_2$, $|2\omega_1 - \omega_2|$—combination frequencies.

Continuing with this analysis for the more higher powers of series (8.8), one can show that when a biharmonic oscillation acts upon a

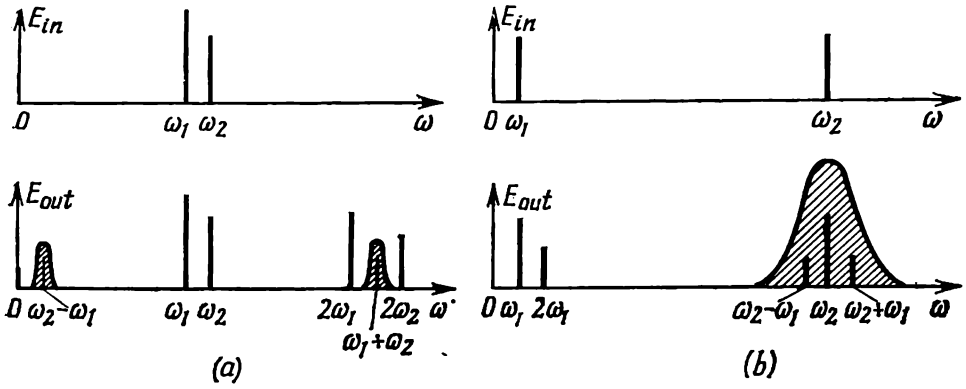


Fig. 8.14. Spectra of an oscillation at the input and output of a square-law nonlinear element in the case of a biharmonic excitation with frequencies ω_1 and ω

(*a*) for close frequencies ω_1 and ω_2; (*b*) for $\omega_1 \ll \omega_2$

nonlinear device described by a kth-degree polynomial, the output spectrum of this device can include the following frequencies: $\omega = 0$ — the d-c component; $\omega = n\omega_1$, $n = 1, 2, \ldots, k$ — harmonics of the frequency ω_1; $\omega = n\omega_2$, $n = 1, 2, \ldots, k$ — harmonics of the frequency ω_2; $\omega = |n\omega_1 \pm m\omega_2|$, $n = 1, 2, \ldots, k$ and $m = 1, 2, \ldots, k$ — combination frequencies (provided that $n + m \leqslant k$).

Spectrograms of oscillations at the input and output of a nonlinear device described by a second-degree polynomial ($k = 2$) is shown in Fig. 8.14. From the figure it is clear that the interaction of two harmonic oscillations of different frequencies in a quadratic nonlinear device results in the development of a difference frequency $|\omega_1 - \omega_2|$ and a sum frequency $\omega_1 + \omega_2$ (besides the harmonics $2\omega_1$ and $2\omega_2$). To make practical use of these new frequencies, it is sufficient to employ a linear selective circuit separating the useful spectrum component at the output of the nonlinear element (Fig. 8.14*a*).

The ability of a quadratic nonlinear element to produce combination frequencies is widely used in radio engineering for shifting the frequency of a signal.

In the case of $\omega_1 \ll \omega_2$ (Fig. 8.14*b*), where the combination frequencies $\omega_2 \pm \omega_1$ are located near the frequency ω_2 and all the three frequencies ω_2, $\omega_2 + \omega_1$, and $\omega_2 - \omega_1$ can be separated by a single common filter, one can obtain a spectrum corresponding to *amplitude modulation* of an oscillation of frequency ω_2 relative to the low frequency ω_1. With a nonlinearity of a higher order ($k > 2$), one can separate any frequency of the form $\omega = |n\omega_1 \pm m\omega_2|$, $n + m \leqslant k$.

In the case of a more complex composition of the input spectrum containing frequencies ω_1, ω_2, ω_3, . . ., at the output of a nonlinear element are produced frequencies $n\omega_1$, $n\omega_2$, $n\omega_3$, . . . and combination frequencies $n\omega_i \pm m\omega_k$, where n and m are any integers, while ω_i and ω_k represent any of the pairs of the input spectrum frequencies.

It can easily be found that with any complex but *periodic* excitation of fundamental frequency ω, at the output of a nonlinear element there also occurs a *periodic process of fundamental frequency* ω. In this case, the spectrum may become enriched only owing to the harmonics of frequencies $n\omega$. This is due to the fact that in the given particular case all the frequencies of the output signal are multiples of the frequency ω and consequently, both the sums of and the differences between any two harmonics of the input spectrum are also multiples of ω.

From the qualitative analysis given above it is clear that a simple resistive nonlinear element in conjunction with a selective linear circuit enables one to effect a number of conversions such as nonlinear reasonance amplification, frequency multiplication, rectification, detection of modulated waves, frequency shifting, and amplitude modulation. These questions are discussed in Secs. 8.4 through 8.12.

The action of harmonic oscillations upon reactive nonlinear elements (capacitance or inductance) is discussed in Ch. 10.

8.4. NONLINEAR RESONANCE AMPLIFICATION

As to its circuit diagram, a nonlinear resonance amplifier does not differ from that discussed in Ch. 5 (Fig. 5.17). The main distinctive feature consists in the operating conditions of the amplifying device. By shifting the operating point on the current-voltage characteristic curve to the left and increasing the input oscillation amplitude the amplifier is made to operate with cutoff of the collector current $i_C(t)$ (in a transistor amplifier) or of the anode current $i_a(t)$ (in a vacuum-tube amplifier). Such operation is illustrated in Fig. 8.10*a*.

In the text below, we shall discuss specific features of nonlinear operation, typical of any type of amplifier. The current $i(t)$ in the output circuit of the amplifier operating with current cutoff is pulsed (Fig. 8.11) and includes, along with the d-c component and the useful first harmonic, a number of higher harmonics, which must be suppressed (filtered out). This is effected by a parallel resonant (tank) circuit tuned to the frequency ω_0 of the input oscillation. In the case of parallel resonance, the equivalent impedance $Z_{eq.r}$ of the tank circuit between the points *1* and *2* is very high, and it is the load impedance of the amplifier. But as to the higher harmonics of the current $i(t)$, the tank can be considered to be short-circuted, provided its Q-factor is sufficiently high. As a result, in spite of the distorted, pulsed shape of the current $i(t)$, the voltage produced across the load circuit, is very close to harmonic.

The main advantage of nonlinear operation is its comparatively high efficiency, by which is understood the ratio of the oscillatory power $P_{\sim} = \frac{1}{2} I_1 V_t$ dissipated in the tank circuit to the power $P_0 = I_0 E_0$ consumed from the d-c current source. Thus

$$\text{Efficiency} = P_{\sim}/P_0 = \frac{1}{2} \frac{I_1}{I_0} \frac{V_t}{E_0}$$

The amplitude of the voltage V_t across the tank circuit can be brought close to E_0, while the current ratio $I_1/I_0 = \gamma_1$ at a cutoff angle $\theta = 80°$ to $100°$ can be made close to $\pi/2$. Therefore, the efficiency of a nonlinear amplifier can be brought to (70 to 80)%, while in linear operation where the amplitude of the a-c component of the current $i(t)$ must be at least several times lower than the quiescent current $i(E_0)$ (see, for example, Fig. 5.12*b*), the efficiency does not exceed a few percent. (In the resonance amplifiers used in radio receivers, the ratio I_1/I_0 is so low that the efficiency is of no consideration at all).

From the graphs in Fig. 8.12 it follows that to increase the ratio $\gamma_1 = I_1/I_0$, it is advantageous to reduce the cutoff angle θ. In this case, however, the value of I_1 is reduced (at a given value of the pulse I_m) and this results in a decrease of the power $P_{\sim}$ (the power P_0 drops down faster than $P_{\sim}$). Therefore, where it is necessary to maximize the power $P_{\sim}$, the cutoff angle θ is brought to about 120°, in which case the coefficient $\alpha_1(\theta)$ reaches its maximum, putting up with some reduction of efficiency. In practical applications of nonlinear amplifiers, the cutoff angle is kept near 90°.

Now let us find the relations between the voltages and currents of the fundamental frequency ω_0 in a nonlinear amplifier.

As a first approximation, if no account is taken of the reverse action of the output circuit on current, i.e., if the current $i(t)$ is supposed to be essentially defined by the voltage at the amplifier

input, one can use formula (8.20) which, with account being taken of (8.26), gives

$$I_m = a_1 E (1 - \cos \theta) = I_1/\alpha_1 (\theta)$$

from which we have

$$I_1 = \alpha_1 (\theta) I_m = \alpha_1 (\theta) (1 - \cos \theta) a_1 E \qquad (8.31)$$

It should be recalled that, in accordance with expression (8.9), the coefficient $a_1 = S$ is the slope of the current-voltage characteristic curve in its linear section.

Thus,

$$I_1 = \alpha_1 (\theta) (1 - \cos \theta) SE \qquad (8.31')$$

The equivalent circuit of the output circuit of the amplifier is shown in Fig. 8.15*a*. The active element is replaced by a pulse-

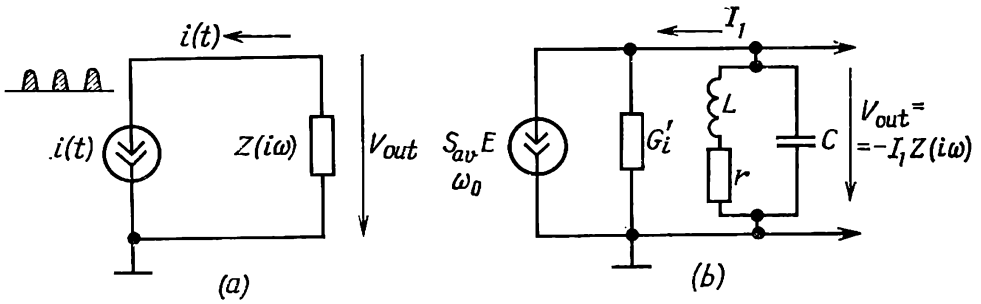


Fig. 8.15. Common equivalent circuit of the output circuit of an amplifier (*a*) for the amplifier operating with current cutoff; (*b*) for the first harmonic of a pulsed current

current generator, but the voltage across the tank circuit is produced by the first harmonic only and therefore, is defined by the expression

$$v_t (t) = -I_1 Z_{eq,\, r} \cos \omega_0 t = V_{out} \cos \omega_0 t \qquad (8.32)$$

(The minus sign accounts for the selected direction of current in the circuit shown in Fig. 8.15 and also for the fact that potentials are read with respect to the earthed point of the circuit).

Dividing expression (8.31) by E, we obtain the parameter

$$S_{av} = a_{1,\, av} = I_1/E = S (1 - \cos \theta) \alpha_1 (\theta) \qquad (8.33)$$

which may be regarded as the average slope of the characteristic curve for the first harmonic.

Thus,

$$I_1 = S_{av} E \qquad (8.34)$$

In contrast to the differential slope $S = a_1$, which is determined at a point and therefore, when using the nonlinear portion of the characteristic curve, depends on the given instant of time, the parameter S_{av}, expressed as the ratio between current and voltage amplitudes, is one averaged, as it were, over the whole period of oscillation.

The concept of average slope makes sense if the voltage across the load is sinusoidal (in spite of the complex shape of the current $i(t)$).

Taking into account the effect of the output voltage on the current $i(t)$, expression (8.34) must be replaced by a more accurate one, similar to expression (5.37′):

$$I_1 = S_{av}E - V_{out}G_i' = S_{av}E - V_{out}/R_i' \qquad (8.34')$$

where

$$G_i' = 1/R_i' = G_i\alpha_1(\theta)(1 - \cos\theta) \qquad (8.35)$$

is the internal conductance of the nonlinear element, *referred to the current of the first harmonic.*

Substituting $I_1 = V_{out}/Z_{eq,\,r}$ into (8.34′) and taking into account (8.32), one can easily obtain the following expression for the amplification factor in operation with current cutoff:

$$K_E = \frac{V_{out}}{E_1} = -\frac{S_{av}Z_{eq,\,r}}{1 + Z_{eq,\,r}/R_i'} \qquad (8.36)$$

For $Z_{eq,\,r}/R_i \ll 1$, one may use the approximate formula

$$K_E \approx -S_{av}Z_{eq,\,r} \qquad (8.36')$$

On the basis of expression (8.34′), the equivalent circuit of the output circuit of the amplifier can be reduced to the form shown in Fig. 8.15*b*. The symbol $\mathbf{V}_{out} = -\mathbf{I}_1 Z(i\omega)$ stands for the complex amplitude of the output voltage.

From a similar equivalent circuit of a linear amplifier (Fig. 5.17*b*) this circuit differs in that here S_{av} and G_i' are functions of the cutoff angle θ and hence, of the amplitude of the input voltage E.

At $\theta = 0$ the amplifying element is cut off completely and $S_{av} = 0$. At $\theta = 90°$, when the current has the shape of half-wave pulses, $S_{av} = a_1/2$, and at $\theta = 180°$ (linear operating conditions) the average slope S_{av} tends to $S = a_1$.

The fact that a change in the amplitude of oscillation results in a change in the parameters S_{av} and G_i' and consequently, in disproportion between the amplitudes at the input and output of the circuit, forces one to treat this circuit as a nonlinear one. On the other hand, the preservance of the shape of oscillation (harmonic) makes it possible to treat the circuit as a linear one (with a fixed amplitude).

Such an approach to the analysis of nonlinear devices is known as the *quasi-linear method* [5]. It should be noted once more that this method is applicable only in cases where, in spite of the nonlinearity of the circuit, the sinusoidal shape of oscillations is maintained, the system being analyzed under steady-state conditions.

The dependence of the parameter S_{av} on the amplitude of the input signal limits the possibilities of using nonlinear operation to amplify an oscillation which carries information in its envelope

(i.e., an amplitude-modulated wave). Operation with current cutoff of exactly 90° is an exception. From Fig. 8.10 it is evident that at $V_0 = V_1$ a change in the amplitude of the input voltage E leads only to a proportional change in the amplitude of the current pulses, the shape of the pulses being maintained. Thus, in operation with cutoff angle $\theta = 90°$, the average slope is independent of the amplitude of the input signal and is always equal to $S/2$. In this case the first-harmonic coefficient $\alpha_1 = I_1/I_m = 0.5$ [see (8.24)], i.e, the amplitude of the first harmonic is half the amplitude of the pulse.

When amplifying a frequency- or a phase-modulated oscillation, the nonlinearity of amplification is not a limiting factor (regardless of the cutoff angle).

These specific features of nonlinear amplification conditions are of great importance and are widely used in practice.

8.5. FREQUENCY MULTIPLICATION

The fact that a pulsed current contains harmonics with frequencies that are multiples of the fundamental exitation frequency makes it possible to use an amplifier operating with current cutoff as a frequency multiplier. For this purpose, no changes are required in the circuit of the resonance amplifier, it being sufficient to tune the load tank circuit to the frequency of the harmonic to be separated and establish such operating conditions of the active element as are most advantageous for the isolation of the useful harmonic. From the graphs shown in Fig. 8.12 it is evident that to double frequency, it is expedient to operate with a cutoff angle close to 60°, at which the coefficient of the second harmonic passes through its maximum, to triple frequency, a cutoff angle of 40° should be used, etc.

If the tank circuit is tuned to a frequency $n\omega_0$, $n = 2, 3, \ldots$, the current harmonics of the order of $n - 1$ and lower pass essentially through the inductive branch, while the harmonics of the order $n + 1$ and higher pass through the capacitive branch. With a sufficiently high Q-factor of the circuit, the voltage of all the harmonics but the nth is very low. Therefore, the voltage across the tank circuit is close to a harmonic voltage of frequency $n\omega_0$.

It should be noted that to utilize the power of the electron device to the full, the reduction of the cutoff angle must be effected while keeping the pulse amplitude at a constant level. To do this, the amplitude of the input a-c voltage E must be increased simultaneously with the bias voltage $|V_0|$. In Figure 8.16, the bias voltage V_{01} corresponds to the angle $\theta = 90°$, the bias voltage V_{02} corrseponds to the angle $\theta = 60°$, etc. The amplitudes E_1, E_2, . . . are selected such that I_m remains unchanged. Therefore, we may say that the opera-

ting conditions of a frequency multiplier are characterized by high amplitudes of the input voltage.

This circumstance and also, the reduction of the useful power with increasing order of multiplication due to the decrease of the coefficients α_n (see Fig. 8.12) substantially impair the energy relations in the multipliers.

The equivalent circuit of a frequency multiplier does not differ from that of a nonlinear amplifier (see Fig. 8.15*b*). In this case, one should, by analogy with expression (8.33), express the average slope as

$$S_{av} = I_n/E = S\,(1 - \cos\theta)\,\alpha_n \qquad (8.37)$$

where the coefficient α_n of the nth harmonic is defined by formula (8.26)

Correspondingly, the internal resistance of the electron device, referred to the selected harmonic, is

$$R_i' = R_i/\alpha_n\,(1 - \cos\theta) \qquad (8.38)$$

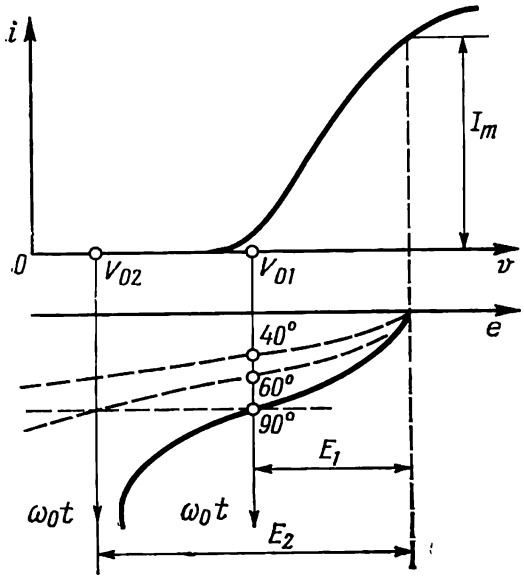


Fig. 8.16. Selecting the cutoff angle in a frequency multiplier for different multiplication factors

Frequency multiplication is widely used in radio transmitters with quartz master oscillators. The frequency of such an oscillator is selected to be relatively low, 4 to 12 times lower than the working frequency of the transmitter, thus providing favourable conditions for utilizing the piezoelectric effect of the quartz bar. Frequency multiplication is effected in the subsequent low-power stages of the transmitter. Most often, in a single stage frequency is doubled or, more seldom, tripled.

Frequency multiplication is also widely used in measuring instruments, where it is necessary to obtain a series of frequencies that are multiples of a certain definite frequency used as a reference one. In such devices, use is made of an electron device which operates with a very small cutoff angle. Applying a sufficiently high a-c voltage to the input (under conditions of a large bias), one can obtain a current in the form of a train of very sharp pulses. Such a current is reach in harmonics that form a wide spectrum. With this spectrum acting upon a tank circuit, the voltage across this circuit may differ considerably from sinusoidal, for a number of harmonics enter into

the passband of the circuit. In such cases, it often proves convenient to determine the voltage across the tank circuit proceeding not from the spectral representation of the pulsed current, but from the consideration of free oscillations excited by each separate current pulse (Fig. 8.17). In the interval T between two current pulses, the amplitude of the voltage across the tank circuit diminishes according to

$$V(t) = V_0 e^{-\alpha t} = V_0 e^{-\omega_{fr} t/2Q}$$

where ω_{fr} is the frequency of free oscillations in the circuit and Q is the quality factor of the circuit.

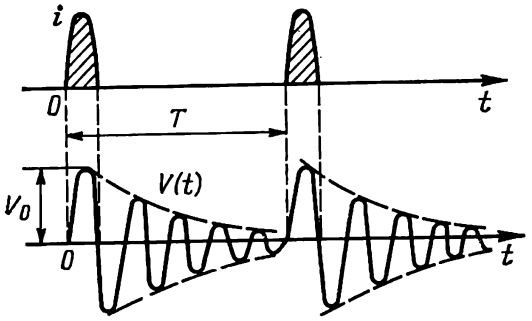


Fig. 8.17. Output voltage of a frequency multiplier in the case of an insufficiently high Q-factor of the resonance circuit

If the oscillation generated by a pulse has not enough time to decay completely by the beginning of the next pulse, the superposition of the free oscillations must be taken into account.

The above picture of phenomena occurring in a frequency multiplier is only a qualitative illustration of an essentially nonlinear operation of an amplifying element and of the separation of the useful harmonic by means of a selective circuit. When designing a frequency multiplier, it is necessary to take into account the deformation of the current pulses due to the nonlinear character of the internal resistances of the amplifying device. This deformation manifests itself in semiconductor devices.

8.6. AMPLITUDE LIMITING

In radio engineering, it often proves necessary to eliminate unwanted changes in the amplitude of an r-f oscillation, which may occur when noise is superimposed on the useful signal, when transmitting frequency-modulated waves through selective networks, etc.

For this purpose, wide use is made of amplitude limiters that are a combination of a nonlinear element and a selective load. The current-voltage characteristic of the nonlinear element must have a pronounced horizontal portion, and the passband of the selective network must not be wider than that required for transmitting the information contained in the frequency (or phase) of the oscillation being limited. An ordinary resonance amplifier, such as that described in Sec. 8.4, operating under the conditions illustrated in Fig. 8.18 may be used as an amplitude limiter.

Let the limiter be fed with an oscillation of the form

$$e(t) = E(t) \cos [\omega_0 t + \theta(t)] \tag{8.39}$$

the change of the envelope $E(t)$ being an undesirable, parasitic factor. If this change does not go beyond the limits of the horizontal

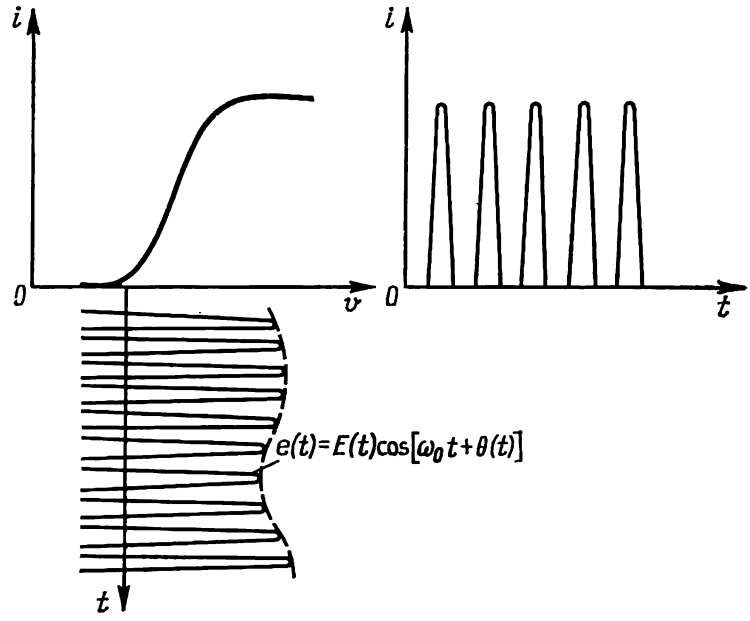


Fig. 8.18. Operating conditions of an amplitude limiter

portion of the characteristic curve $i = f(v)$, as shown in Fig. 8.18, all the current pulses have the same amplitude regardless of $E(t)$. In this case, it is only the width of the top portion of the pulses that changes to some extent. Therefore, as a first approximation, we may consider that the amplitude of the first harmonic and hence, the amplitude of the voltage across the tank circuit is *constant* within some range of the amplitude $E(t)$.

The characteristic curve of a limiter with a selective load suppressing the higher harmonics may be represented in the form shown in Fig. 8.19. In this figure, E_{thr} stands for the threshold value of the amplitude of the input voltage, starting from which complete limiting at the V_0 level is ensured.

When $E(t) > E_{thr}$ the amplitude of the output voltage remains almost unchanged. The phase of the first harmonic and correspondingly, the phase of the output voltage coincides with that of the voltage at the limiter input.

Therefore, we may write the following expression for the output voltage:

$$v_{out}(t) \approx V_0 \cos[\omega_0 t + \theta(t)] \tag{8.40}$$

The amplitude V_0 of the output voltage is defined by the parameters of the nonlinear element and selective load. For the circuit shown in Fig. 8.15*b*, $V_0 = I_1 Z_{eq,r}$, where I_1 is the amplitude of the first harmonic, determined with due regard for the flattening of the pulse top, and $Z_{eq,r}$ is the equivalent impedance of the tank circuit at resonance.

The action of two signals with close frequencies upon an amplitude limiter is of special practical importance.

For example, let the voltage $e(t)$ determined by expression (8.39) be a sum of two harmonic oscillations:

$$e(t) = E_1 \cos \omega_1 t + E_2 \cos \omega_2 t, \quad E_2 < E_1 \tag{8.41}$$

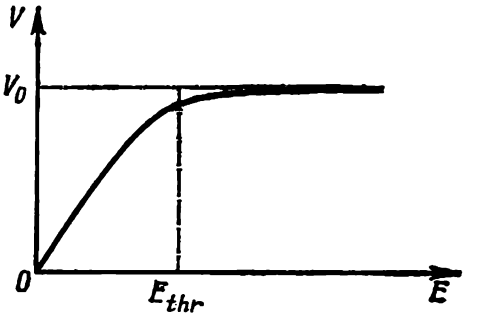


Fig. 8.19. Characteristic curve of a resonance limiter

Each of these voltages, while acting separately, produces at the limiter output a simple harmonic oscillation with a frequency ω_1 (or ω_2) and an amplitude V_0. A different picture is observed when the two harmonic voltages act upon the limiter simultaneously. To find the output voltage of the limiter, it is necessary to reduce the input voltage to the form (8.39).

For this purpose, let us denote the frequency difference $\omega_2 - \omega_1$ by Ω and substitute into (8.41) the following expression:

$$\cos \omega_2 t = \cos(\omega_1 + \Omega)t = \cos \Omega t \cos \omega_1 t - \sin \Omega t \sin \omega_1 t$$

Then

$$e(t) = E_1 \cos \omega_1 t + E_2 (\cos \Omega t \cos \omega_1 t - \sin \Omega t \sin \omega_1 t)$$

$$= (E_1 + E_2 \cos \Omega t) \cos \omega_1 t - E_2 \sin \Omega t \sin \omega_1 t$$

Considering the factors before $\cos \omega_1 t$ and $\sin \omega_1 t$ to be slowly varying functions of time (since $\Omega \ll \omega_1$), let us write the last expression in a somewhat modified form:

$$e(t) = \sqrt{(E_1 + E_2 \cos \Omega t)^2 + E_2^2 \sin^2 \Omega t} \cos[\omega_1 t + \theta(t)]$$

$$= E(t) \cos[\omega_1 t + \theta(t)] \tag{8.42}$$

where the envelope $E(t)$ of the resultant voltage is determined by the expression

$$E(t) = E_1 \sqrt{1 + (2E_2/E_1) \cos \Omega t + (E_2/E_1)^2} \tag{8.43}$$

and the phase is given by

$$\theta(t) = \arctan \frac{(E_2/E_1)\sin \Omega t}{1+(E_2/E_1)\cos \Omega t} \tag{8.44}$$

The envelope $E(t)$ has a maximum value equal to $E_1 + E_2$ (for $\cos \Omega t = 1$) and a minimum value equal to $E_1 - E_2$ (for $\cos \Omega t = -1$).

Let $E_1 - E_2 > E_{thr}$ so that the condition for limitation is satisfied for all values the amplitude $E(t)$ of the input voltage can have

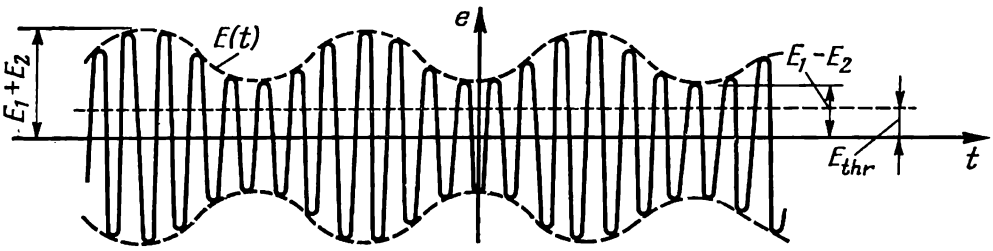


Fig. 8.20. Biharmonic voltage at the input of an amplitude limiter

(Fig. 8.20). Then, by analogy with (8.40), the output voltage may be written as

$$v_{out}(t) = V_0 \cos[\omega_1 t + \theta(t)] \tag{8.45}$$

Thus we obtain a phase-modulated oscillation which, in contrast to the input voltage $e(t)$, may have a wide spectrum.

To determine the amplitudes of the individual components of this spectrum, one can use the theory of frequency-modulated oscillations given in Ch. 3.

Without going into a detailed analysis, let us simplify the problem by setting $E_2 \ll E_1$. In this case, expression (8.44) is reduced to

$$\theta(t) \approx \arctan\left(\frac{E_2}{E_1}\sin \Omega t\right) \approx \frac{E_2}{E_1}\sin \Omega t$$

and the output voltage is given by*

$$v_{out}(t) \approx V_0 \cos(\omega_1 t + m \sin \Omega t) \tag{8.46}$$

Here we use the symbol

$$m = E_2/E_1 \ll 1 \tag{8.47}$$

which emphasises that in this case the ratio of the amplitudes E_2/E_1 has the meaning of the phase-modulation index.

Expression (8.46) completely coincides with (3.25) and, by analogy with (3.32), may be written in the form

$$v_{out}(t) = V_0[\cos \omega_1 t + (m/2)\cos(\omega_1 + \Omega)t - (m/2)\cos(\omega_1 - \Omega)t] \tag{8.48}$$

* Here we do not consider the effect of the nonuniformity of the frequency characteristic of the tank circuit in the frequency band resulting from the phase modulation of the input signal.

With $m = E_2/E_1 \ll 1$, the output voltage spectrum consists of three components with frequencies ω_1, $\omega_1 + \Omega = \omega_2$ and $\omega_1 - \Omega = 2\omega_1 - \omega_2$. The first two frequencies are present at the limiter input, while the third frequency $(2\omega_1 - \omega_2)$ is the product of interaction of the input oscillations in the nonlinear element.

The relation between the spectra at the input and output of the limiter for the case $E_2/E_1 \ll 1$ is shown in Fig. 8.21. The frequency $2\omega_1 - \omega_2$ is a "mirror" frequency with respect to ω_2.

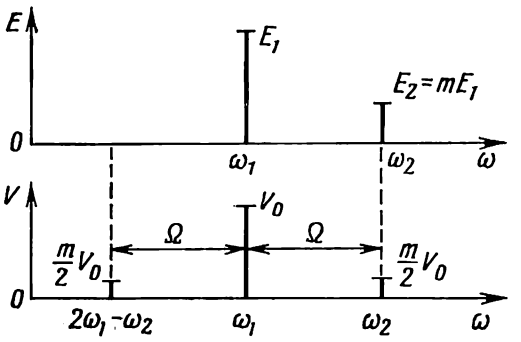


Fig. 8.21. Oscillation spectra at the input and output of a resonance limiter in the case of a biharmonic excitation

It is essential that the amplitude of the oscillation with a frequency $\omega_2 = \omega_1 + \Omega$ (as well as of that with a frequency $\omega_1 - \Omega$) is only $m/2$ of the amplitude of the oscillation with a frequency ω_1, while at the input the amplitude ratio is equal to m. We may say that in the limiter the weak oscillation is being suppressed by the stronger one. The suppression effect is particularly evident where the passband of the selective load is such that only the frequencies ω_1 and ω_2 are let pass through, while the mirror frequency $\omega_1 - \Omega$ is rejected.* In this case the spectrum of the output voltage is the same as that of the input voltage, only the amplitude of the weak oscillation is reduced two times with respect to the strong oscillation.

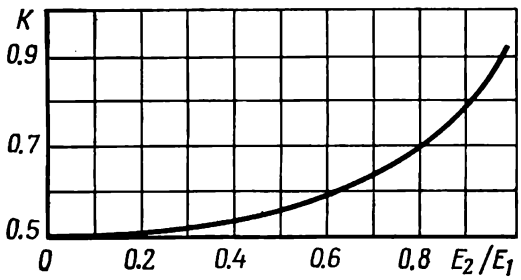


Fig. 8.22. Rejection ratio for an amplitude limiter in the case of a biharmonic excitation $(E_2 < E_1)$

If the amplitudes of the two oscillations at the limiter input are commensurable, the effect of relative suppression is less pronounced. This is clear from the graph in Fig. 8.22 plotted under the assumption that the selective network of the limiter lets pass through only the components of frequencies ω_1 and ω_2 that are present at the limiter input [8]. Plotted along the ordinate axis is the rejection ratio, that

* This, however, is not typical of the limiter, for the rejection of the mirror frequency causes fluctuation of the envelope of the sum of the two voltages.

is the ratio

$$K = \frac{V_2/V_1}{E_2/E_1}$$

where E_2/E_1 is the ratio of the amplitudes at the input and V_2/V_1, that at the output of the limiter.

When $E_2/E_1 \ll 1$, $K = 0.5$, but as E_2/E_1 approaches unity, K also approaches unity. When $E_2/E_1 = 1$, both oscillations are equivalent and there is no mutual suppression.

In conclusion, it should be noted that all the above arguments are also valid for the case $\omega_2 < \omega_1$, only the mirror frequencies in Fig. 8.21 should be interchanged.

8.7. NONLINEAR RECTIFYING CIRCUIT

Let us consider the nonlinear circuit shown in Fig. 8.23. Here, a harmonic e.m.f. $e(t) = E \cos \omega_0 t$ is applied across a nonlinear element D (a diode) and a simple R-C filter connected in series. It is required to find the currents in the branches and the output volta-

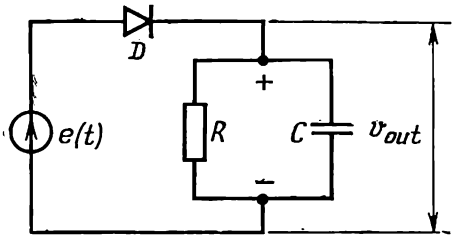


Fig. 8.23. Half-wave rectifier

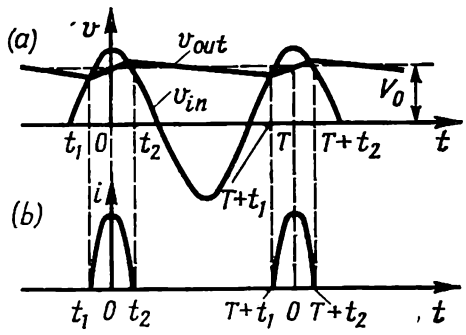


Fig. 8.24. (*a*) Voltages at the input and output of a half-wave rectifier and (*b*) the current in the diode circuit

ge v_{out} of the circuit (under steady-state conditions). Such a problem is typical of half-wave rectification of an a-c current, amplitude detection (in the absence of modulation), and many other radio engineering processes. The output voltage $v_{out}(t)$ is represented by a curve pulsing about some average value V_0 (Fig. 8.24*a*). This voltage is negative with respect to the diode. Therefore, the current through the diode can flow only when the positive half wave of the e.m.f. exceeds the voltage $v_{out}(t)$. In other words, the current flowing through the diode has the form of pulses shown in Fig. 8.24*b*. Between the pulses, while the capacitor C discharges through the resistor R, the voltage $v_{out}(t)$ decreases. During the time interval

$t_1 < t < t_2$, the capacitor is charged up by a current pulse and $v_{out}(t)$ rises. If the time constant of the R-C load circuit is large compared to the period $T = 2\pi/\omega_0$, the amplitude of the pulsations of the voltage v_{out} is low and as a first approximation, we may consider that $v_{out} \approx V_0$. Taking into account that the voltage across the load is negative with respect to the diode, let us study the plot shown in Fig. 8.25. The continuous line on the left-hand side of this

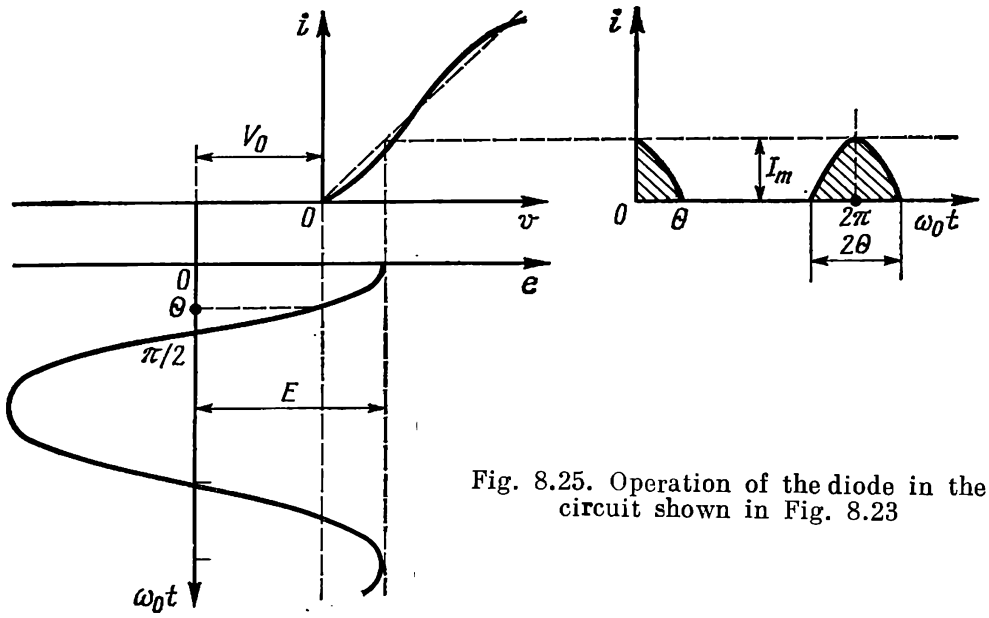


Fig. 8.25. Operation of the diode in the circuit shown in Fig. 8.23

figure shows the actual current-voltage characteristic of the diode in coordinates i, v, while the dashed line shows a linear function approximating this characteristic. The diagram of the input e.m.f. $e(t) = E \cos \omega_0 t$ is plotted with respect to the vertical axis $\omega_0 t$ shifted to the left from the point $v = 0$ by an amount of V_0. Shown on the right-hand side of Fig. 8.25 are the current pulses whose duration is equal to 2θ.

In contrast to the plot shown in Fig. 8.10, here we have no fixed d-c voltage. It should be borne in mind that the d-c voltage V_0 produced across the load resistor R by the d-c component of the current I_0 depends on the amplitude E of the input oscillation. In particular, from this it follows that *the cutoff angle* θ *cannot exceed* 90°.

Let us use the results of the spectral analysis of a pulsed current, given in Sec. 8.3, for establishing the relation between the amplitude of the input voltage E and the rectified voltage V_0, the parameters of the circuit being specified.

First, assume that the current cutoff angle θ is known. In this case, we may write the following relations:

$$I_0 = \alpha_0(\theta)\, I_m \tag{8.49}$$

$$\cos\theta = V_0/E \tag{8.50}$$

The last relation stems directly from Fig. 8.25. Then, with the internal resistance R_i of the diode being specified, we have

$$I_m = (E - V_0)/R_i = E\,(1 - V_0/E)/R_i \tag{8.51}$$

Substituting (8.49) and (8.50) into this expression, we obtain

$$\frac{I_0}{\alpha_0(\theta)} = \frac{E(1-\cos\theta)}{R_i} = \frac{V_0}{\cos\theta}\,\frac{(1-\cos\theta)}{R_i}$$

whence

$$\frac{I_0}{V_0} = \frac{1}{R} = \frac{\alpha_0(1-\cos\theta)}{\cos\theta}\,\frac{1}{R_i}$$

and finally, taking into account the first equality of (8.26),

$$\frac{R_i}{R} = \frac{\sin\theta - \theta\cos\theta}{\pi\cos\theta} = \frac{\tan\theta - \theta}{\pi} \tag{8.52}$$

Thus, specifying the internal resistance R_i of the diode and the load resistance R unambiguously defines the cutoff angle θ, it being presupposed that the capacitance C shunting the resistance R satisfies the condition

$$1/\omega_0 C \ll R \tag{8.53}$$

or, which is the same, the time constant RC is large compared to the period T_0, for in this case only the output voltage may be considered to be close to d-c voltage.

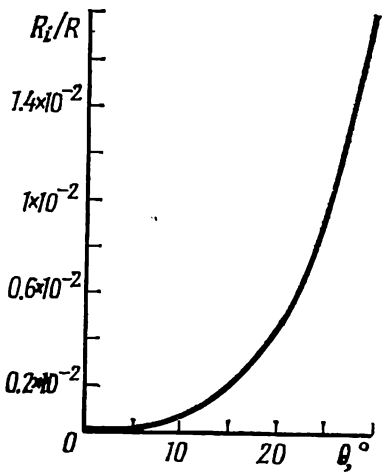


Fig. 8.26. Characteristic R_i/R as a function of the cutoff angle θ

Equation (8.52) relating the cutoff angle θ to the ratio R_i/R is transcendental. Therefore, it is convenient to find θ by the graph representing the dependence of the ratio R_i/R on θ (Fig. 8.26). Let us consider two limiting cases: (1) $\theta = 0$ and (2) $\theta = 90°$. The first case occurs with $R_i/R \to 0$, i.e., with infinitely high load resistance R, when the detector circuit degenerates into the circuit shown in Fig. 8.27. In this case, the rectified voltage across the capacitor C reaches the highest possible value $V_0 = E$ and the current through the diode under steady-state conditions, when the capacitor C has charged completely, is zero. Thus, the case $\theta = 0$ corresponds to no-load conditions. The second case ($\theta = 90°$) corresponds to short-circuit conditions ($R \to 0$). In this

case, the entire e.m.f. is applied across the diode and the diode current assumes the form of half-wave pulses (truncated at the top if E is higher than the diode saturation voltage).

If the effect of the capacitance is disregarded, which is admissible with low R, one will come to the circuit shown in Fig. 8.28. In this

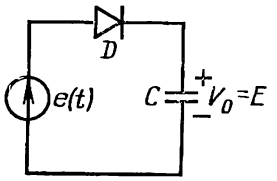


Fig. 8.27. Equivalent circuit of a rectifier operating under no-load conditions ($R_i \to \infty, \theta \to 0$)

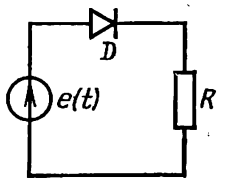


Fig. 8.28. Equivalent circuit of a rectifier for $R \ll R_i$, $\theta \to 90°$

case, the voltage across the resistor R has the same shape as the current i.

Thus, to obtain at the output a rectified voltage with an amplitude close to that of the e.m.f. E, the cutoff angle must be small, while the ratio R/R_i must be very high. At $\theta \ll 10$ to 20°, the ratio $V_0/E = \cos\theta$ is close to unity. To obtain such operating conditions, it is necessary that the load resistance $R \approx 100\ R_i$. After R has been found, the required capacitance of the capacitor C can be calculated using conditions (8.53).

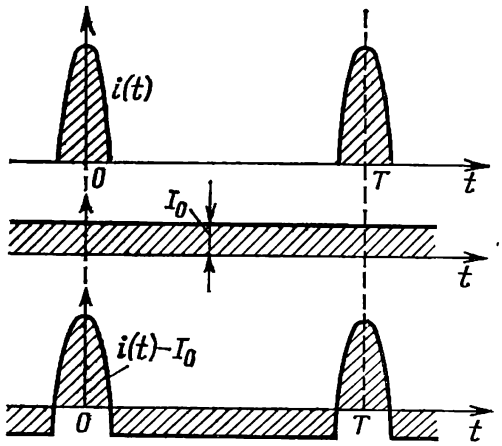


Fig. 8.29. Pulsed current in a diode circuit and its components

In conclusion, it should be noted that condition (8.53), derived by considering the process of discharge of the capacitor C through the resistor R, can easily be interpreted on the basis of spectral approach. When $1/\omega_0 C \ll R$, all the harmonics of the pulsed current flowing through the diode are shorted mainly through the capacitor, without producing across it any noticeable voltage drop (as compared with $V_0 = I_0 R$). As a result, the current distribution will have the form shown in Fig. 8.29. The current shown in the lower part of the figure and obtained by subtracting the d-c component I_0 from the total diode current $i(t)$, is the sum of all the harmonics of this current.

8.8. AMPLITUDE DETECTION

The detection of oscillations is the separation of a signal from a modulated r-f wave which contains it in an implicit form. This process is opposite to modulation, so where it is necessary to emphasize this fact, the term "detection" is replaced by the term "demodulation". By analogy with the basic types of modulation, distinction is made between amplitude, frequency, and phase detection. The last two types of detection are effected by means of almost identical devices, for the frequency and phase of an oscillation are closely interrelated.

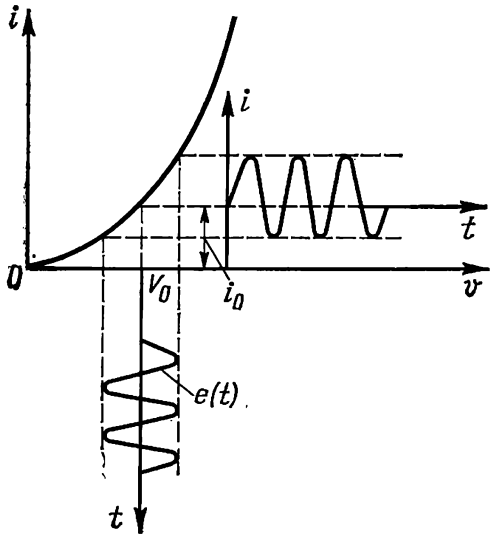


Fig. 8.30. Operation of a square-law detector

The input of the detector is fed with a modulated oscillation including radio-frequency components only, i.e., the carrier and the side-frequency oscillations, while the output voltage includes the low-frequency spectrum of the message being transmitted. Consequently, detection is accompanied by conversion of frequency spectrum and can be effected only by using nonlinear circuits or linear circuits with variable parameters. At present, semiconductor diodes are predominantly used as nonlinear elements for detection.

The principle of operation of the amplitude detector in the absence of modulation was described in the preceding section when the process of rectification was discussed. Now we shall discuss some phenomena taking place in the detector supplied with a modulated oscillation and also, some specific features of detection of weak and strong signals. Let us first discuss the latter question. Assume that the amplitude of the signal at the detector input is so low that the current changes caused thereby are restricted to a comparatively small section of the lower bend of the characteristic curve of a diode or any other nonlinear element (Fig. 8.30).

In accordance with expression (8.10), the current through the diode (see Fig. 8.23) is

$$i(t) = i(V_0) + a_1 e(t) + a_2 e^2(t)$$

where $e(t) = E(t) \cos \omega_0 t$ is the instantaneous value of the r-f signal whose amplitude $E(t)$ is modulated according to the message

being transmitted (the epoch angle is omitted, since it has no effect on the operation of the amplitude detector).

Thus, we have

$$i(t) = i(V_0) + a_1 E(t)\cos\omega_0 t + a_2 E^2(t)\cos^2\omega_0 t$$

$$= i_0 + a_1 E(t)\cos\omega_0 t + \frac{1}{2}a_2 E^2(t)\cos 2\omega_0 t + \frac{1}{2}a_2 E^2(t) \qquad (8.54)$$

The d-c component of the current $i(V_0) = i_0$ (quiescent current) and the r-f components ω_0 and $2\omega_0$ are rejected in the load circuit.

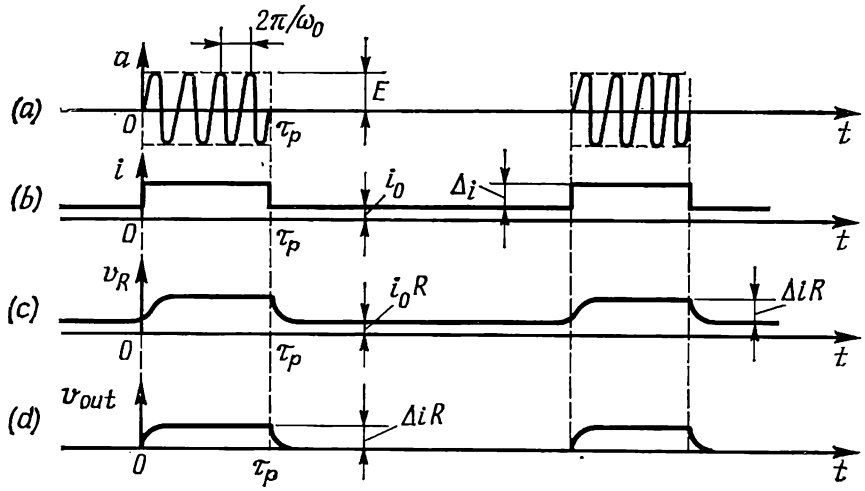


Fig. 8.31. (*a*) Voltage at the input of a square-law detector, (*b*) the d-c component in the diode circuit, (*c*) the voltage across the resistor R, and (*d*) the voltage increment due to the action of the input signal

The information is contained in the last low-frequency term

$$i_{lf} = \frac{1}{2}a_2 E^2(t) \qquad (8.55)$$

Since this component is proportional to the square of the input voltage amplitude, the *detection of low amplitudes is a square-law process.* This general principle holds true for any types of nonlinear element used for detection.

The fact that the voltage $v_{out}(t)$ across the load (which is a linear circuit) is proportional to the l-f current i_{lf} and hence, to the square of the amplitude $E(t)$ of the input signal does not affect the correct reproduction of the shape of square-pulse signals. For example, let the voltage at the detector input have the form of high-frequency pulses with a rectangular envelope (Fig. 8.31*a*). The average diode current in the intervals between the pulses (Fig. 8.31*b*) is the same as the quiescent current i_0, while in the presence of pulses it differs

from the latter by the amount

$$\Delta i = \frac{1}{2} a_2 E^2$$

where E is the amplitude of the r-f voltage, which is constant within the duration τ_p of the pulse.

The voltage $v_R(t)$ across the load of the detector is shown in Fig. 8.31*c*. When the process of charging or discharging of the capacitor C is over, the voltage across the load is equal to $i_0 R$ (in the interval between the pulses) or $(i_0 + \Delta i) R$ (in the presence of a signal). Figure 8.31*d* shows separately the voltage increment caused by the signal. To isolate this increment from the d-c voltage $i_0 R$ one can employ a separation circuit consisting of a capacitor and a resistor.

The output voltage $v_{out}(t)$ shown in Fig. 8.31*d* has a shape similar to that of the envelope of the r-f voltage applied to the detector input. Thus, we make certain that the square law of detection does not impede the reproduction of square pulses. In this case, the nonlinearity of detection characteristic manifests itself only in the fact that the *pulse amplitude at the detector output is proportional to the square of the amplitude of the r-f voltage applied to the detector input.*

This is not the case with the square-law detection of oscillations whose envelope is a continuous function of time which takes place, for example, when transmitting speech, music, etc. To simplify discussion, let us consider the case of tone modulation. Substituting

$$E(t) = E_0 (1 + M \sin \Omega t)$$

into expression (8.55), we obtain

$$i_{lf} = \frac{a_2}{2} E_0^2 (1 + M \sin \Omega t)^2 = \frac{a_2}{2} E_0^2 (1 + 2M \sin \Omega t + M^2 \sin^2 \Omega t)$$

$$= \frac{a_2}{2} E_0^2 \left(1 + \frac{M^2}{2} + 2M \sin \Omega t - \frac{M^2}{2} \cos 2\Omega t\right)$$

It should be noted that in the absence of modulation ($M = 0$), i.e., when the detector is acted upon by the carrier-frequency oscillation only, the current increment is equal to $(a_2/2)E_0^2$. Thus, in the process of tone modulation the average value of current gets a constant relative increment equal to $M^2/2$. The alternating component of current includes the following two terms: (a) the useful component $2M \sin \Omega t$ reproducing the signal and (b) the parasitic component $(M^2/2) \cos 2\Omega t$, which is the second harmonic of the signal.

From this it follows that the harmonic content, which in this case is equal to the ratio of the amplitude of the second harmonic to that of the first one, is

$$K_{h2} = 0.5\, M^2/2M = M/4$$

For 100-percent modulation, we have

$$K_{h2} = 0.25 = 25\%$$

When the signal is simultaneously modulated by two frequencies Ω_1 and Ω_2, the output voltage of the detector contains, in addition to the harmonics $2\Omega_1$ and $2\Omega_2$, two combination frequencies $\Omega_1 + \Omega_2$ and $\Omega_1 - \Omega_2$ whose amplitudes are proportional to the product of the partial modulation factors M_1 and M_2. This result can easily be obtained if we substitute

$$E(t) = E_0 (1 + M_1 \sin \Omega_1 t + M_2 \sin \Omega_2 t)$$

into (8.54).

When transmitting compound signals which include a large number of frequencies, and a high modulation factor is used, the harmonics

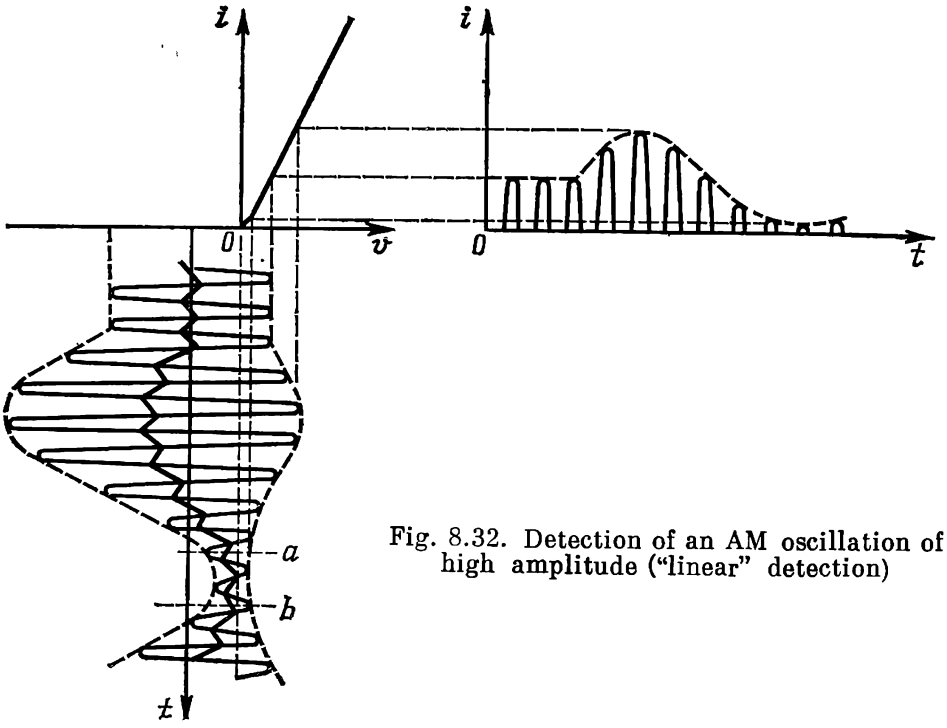


Fig. 8.32. Detection of an AM oscillation of high amplitude ("linear" detection)

and combination frequencies have a very strong effect on the intelligibility and tone of the signal. Therefore the use of the square-law detection is inexpedient where undistorted reproduction of signals is required (when transmitting speech, music, etc.).

Let us consider the detection of high amplitudes. As previously, we shall use a diode detector. Without modifying the system shown in Fig. 8.23, we assume that the amplitude of the input signal is sufficiently high, and that R and C are chosen so that the cutoff

angle is very small and the rectified voltage across R is almost the same as the amplitude $E(t)$ of the input signal. Such conditions for a constant amplitude (rectification) were discussed in Sec. 8.7. In the presence of modulation, the operating conditions of the diode are as shown in Fig. 8.32. The bias voltage produced by the d-c current component varies in proportion to the amplitude of the input signal. But the variable bias voltage is none other than the

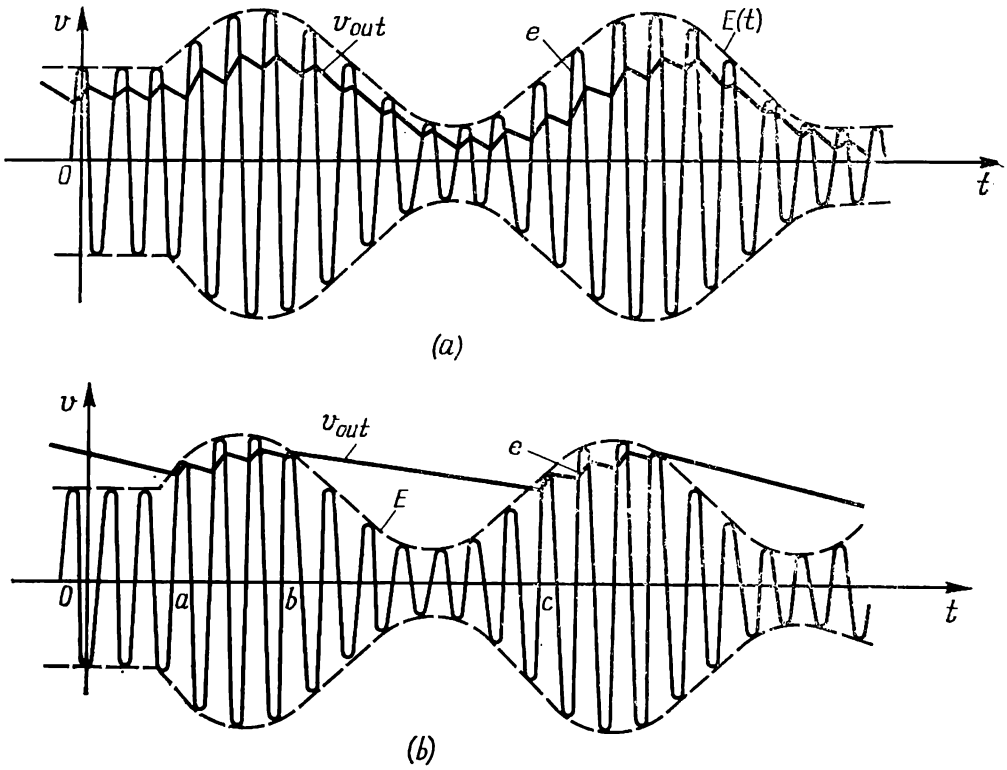


Fig. 8.33 Diagrams of the input and output voltages in a "linear" detector with (*a*) correct and (*b*) incorrect selection of the elements of the *R-C* load circuit

output voltage of the detector. Shown superimposed in Fig. 8.33*a* are the r-f input voltage and the rectified output voltage (toothed line). Since with a sufficiently large time constant (compared to the period $T_0 = 2\pi/\omega_0$ of the carrier) the teeth are practically absent, the output voltage reproduces the envelope of the input voltage, i.e., the message being transmitted. Thus, the relation between the output (rectified) voltage $v_{out}(t)$ and the amplitude of the input e.m.f. $E(t)$ is almost linear. In this sense, a detector operating with high input amplitudes and with a load providing for a close agreement between the voltages $v_{out}(t)$ and $E(t)$ is called *linear detector*. Of course, it should be borne in mind that a detector operating

with current cutoff is an *essentially nonlinear device.* This nonlinearity is conditioned by the shape of the characteristic curve not only in the range $v > 0$ (where the characteristic may be close to linear) but also within the whole range of voltages acting on the diode. When operating with current cutoff, the characteristic curve of the diode is a broken line consisting of a section of the abscissa axis (where $v < 0$) and an inclined line (where $v > 0$), with the bend near the point $v = 0$.

The choice of the detector load elements also depends on the mode of modulation. It is required that the time constant of the load circuit be small as compared with the period of modulation, otherwise the variation of the rectified voltage across the load may lag behind the variation of the envelope of the input e.m.f. Such conditions are shown in Fig. 8.33*b*. In the interval *a-b*, the voltage v_{out} lags behind the envelope of the input e.m.f., because of the excessive delay introduced by the *R-C* circuit. At the point *b*, where v_{out} and the amplitude of the modulated e.m.f. became level, the current through the diode and the increase in v_{out} cease. In the interval *b-c*, the source of the e.m.f. and the diode have no effect on the load circuit, and the capacitor *C* discharges through the resistor *R*. Thus, in the interval *b-c*, the voltage follows an exponential curve. In this case, we have *nonlinear distortion of the signal.* Since this distortion is caused by the interaction of the nonlinear element (diode) with the linear circuit (*R-C*), the degree of nonlinear distortion depends not only on the system parameters and depth of modulation, but on the modulation frequency as well. This distortion increases with the frequency and the depth of modulation of the input e.m.f. This distortion can be eliminated by making $RC \ll 2\pi/\Omega$. On the other hand, to smooth out the high-frequency pulsations, it is necessary that $RC \gg 2\pi/\omega_0$ [see (8.53)]. Combining these two conditions, we obtain

$$2\pi/\omega_0 \ll RC \ll 2\pi/\Omega \tag{8.56}$$

As a rule, the frequencies ω_0 and Ω differ greatly, so that condition (8.56) is easily satisfied.

When the envelope is pulse-modulated, it is necessary to substitute on the right-hand side of inequality (8.56) the pulse duration for the modulation period $2\pi/\Omega$, it being supposed that the intervals between the pulses are large compared to the pulse duration. With very short pulses whose duration is only a few times longer than the period $T_0 = 2\pi/\omega_0$ ("high-frequency hunger"), it is difficult to separate the envelope from the r-f carrier.

Having revealed the mechanism of separation of the envelope of a modulated oscillation, let us consider the *detection characteristic*, i.e., the dependence of $v_{out}(t)$ on the amplitude $E(t)$ of the r-f carrier.

In the absence of modulation, when the operation of the detector does not differ in any way from rectification of a radio-frequency oscillation with a constant amplitude E, the relation between v_{out} and E is defined by expression (8.50), i.e., $V_0 = E \cos \theta = \text{const}$. In Sec. 8.7 it was mentioned that the cutoff angle θ in a rectifier is very small, so that V_0/E is very close to unity.

In the presence of modulation, the relation between $v_{out}(t)$ and $E(t)$ does not remain constant. With upward modulation, the cutoff angle decreases still more and the voltage $v_{out}(t) \to E(t)$. With downward modulation, the discrepancy between $v_{out}(t)$ and $E(t)$, on the contrary, increases. With the modulation depth close to 100%, when the amplitude $E(t)$ almost vanishes (section a-b in Fig. 8.32), rectification takes place at the lower bend of the current-voltage characteristic. In this section, the characteristic curve is close to a parabola and detection is quadratic. As a result, the detection characteristic assumes the form shown in Fig. 8.34 (continuous line). *With low amplitudes, modulation is quadratic, while with high amplitudes, it is linear.* The higher the amplitude of the input oscillation,

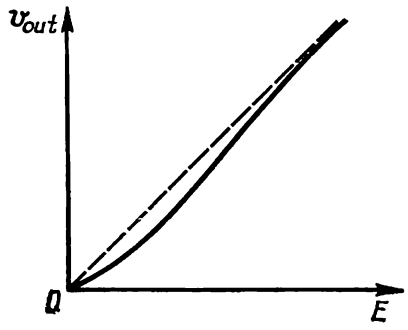


Fig. 8.34. Characteristic of detection of an AM oscillation

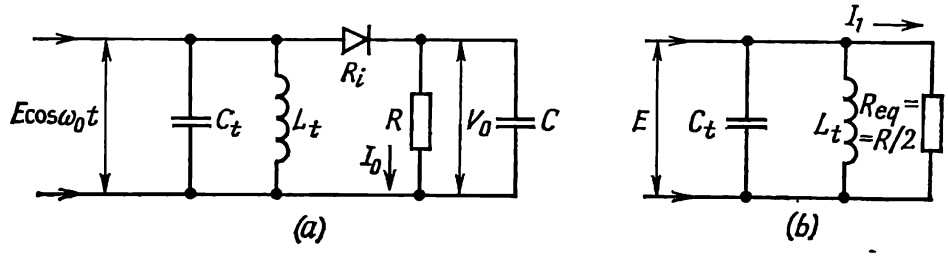


Fig. 8.35. (a) Connection of a diode detector to the tank circuit of an amplifier and (b) the equivalent detector circuit making it possible to define the input resistance of the detector at a frequency ω_0

that corresponds to the peak of modulation, the smaller the effect of deviation of the detection characteristic from a straight line (dashed) near zero.

In conclusion, let us consider the input resistance of a diode detector, i.e., the resistance of a circuit consisting of a diode and a load (R-C) connected in series. This question is of primary importance for determining the attenuation introduced by the detector into the tank circuit of the voltage source (Fig. 8.35a). We shall restrict

ourselves to the case $R \gg R_i$, where the angle θ is so small that we may take $\cos\theta \approx 1$ and $E \approx V_0$.

The power consumed by the detector from the source is equal to $EI_1/2$, where I_1 is the amplitude of the first harmonic of the current flowing through the diode, while the power dissipated in the load resistor is $V_0 I_0$. With $R \gg R_i$, practically all the power consumed by the detector is dissipated in R. Therefore, we may assume that

$$EI_1/2 \approx V_0 I_0$$

Dividing both sides by E^2, we obtain

$$I_1/2E \approx (V_0/E)\,(I_0/E)$$

but

$$V_0/E \approx 1, \quad I_0/E = 1/R, \quad \text{and} \quad I_1 E = 1/R_{eq}$$

where R_{eq} is the sought input resistance of the detector.

From this we find

$$R_{eq} \approx \frac{1}{2} R \tag{8.57}$$

The equivalent circuit of the detector for the frequency ω_0 of the first harmonic I_1 is shown in Fig. 8.35*b*.

The basic results of the study of the principle of amplitude detection by means of a diode can be applied to any other nonlinear elements possessing unilateral conductivity (valve property).

8.9. FREQUENCY AND PHASE DETECTION

The output voltage of a frequency detector must reproduce the modulating function, i.e., the law governing the variation of the *instantaneous frequency* of the modulated wave. Representing the latter in the form

$$e(t) = E(t) \cos [\omega_0 t + \theta(t)] \tag{8.58}$$

we obtain, for an ideal frequency detector, the following relation:

$$v_{out}(t) = S_{fd} \frac{d\theta}{dt} = S_{fd} \Delta\omega(t) \tag{8.59}$$

where $S_{fd} = \text{const}$ is the slope of the characteristic curve of the frequency detector in volts per unit angular frequency and $\Delta\omega(t) = d\theta/dt$ is the instantaneous frequency deviation of the input e.m.f. If frequencies $f = \omega/2\pi$ are used, then the slope S_{fd} of the characteristic curve in the expression

$$v_{out}(t) = S_{fd} \Delta f(t) \tag{8.60}$$

will have the dimension of V/Hz.

It is supposed that $\Delta f(t)$ and, hence, $v_{out}(t)$ are "slow" functions of time. To recover a message from a frequency-modulated oscillation whose spectrum contains only r-f components (the carrier fre-

quency ω_0 and the side frequencies), a *nonlinear device* is needed. Consequently, a frequency detector must include a nonlinear element without fail. However, in contrast to an amplitude detector, in this case a single nonlinear element will not suffice. In fact, from the study of the current-voltage characteristics of nonlinear elements it is clear that the nonlinear element does not respond to a frequency variation of the input voltage having a constant amplitude. In other words, the nonlinearity of such devices as a diode manifests itself only when there is a variation of the operating voltage and not a variation of the frequency or, generally, of the rate of change of the signal. Therefore, a common frequency detector is a combination of the following two basic parts: (1) a linear selective circuit that converts frequency modulation into amplitude modulation and (2) an amplitude detector.

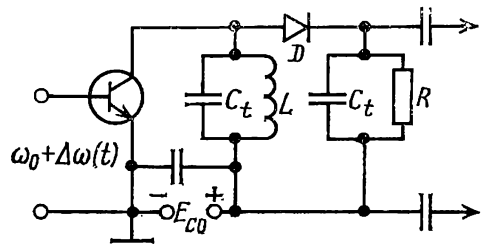


Fig. 8.36. Single-tuned frequency detector

With the detector circuit designed correctly, a change in the amplitude of the input signal has no effect on the output voltage. For this reason, the frequency detector usually includes a circuit to limit the input e.m.f. The limiting is sometimes effected by establishing special operating conditions for the amplifying device included in the frequency detector circuit.

Any electric circuit having a nonuniform frequency characteristic (*R-L* circuits, *R-C* filters, tank circuits, etc.) can be used as linear systems. Oscillatory circuits are widely used in r-f technology. A schematic diagram of a frequency detector comprising a simple tuned (tank) circuit is shown in Fig. 8.36. If the resonance frequency ω_r of the circuit differs from the centre frequency ω_0 of the modulated oscillation, the variation of the amplitude of the voltage V_t across the tuned circuit follows to some extent the variation of the frequency of the input voltage (Fig. 8.37).

A change in the amplitude V_t of the r-f voltage is transformed by means of a diode *D* into a low-frequency voltage across an aperiodic *R-C* load. It should be noted that with exact tuning of the tank circuit to the frequency $\omega_r = \omega_0$ the signal is distorted: the frequency of change of the envelope is twice as high as the frequency of the useful modulation. Under initial conditions, i.e. in the absence of modulation, the operating point must lie on the falling portion of the resonance curve.

The disadvantage of the above scheme consists in that the tank circuit needs to be tuned to a frequency differing from that of the unmodulated oscillation. Furthermore, a solitary tank circuit has

a very limited linear section on the falling portion of the resonance curve.

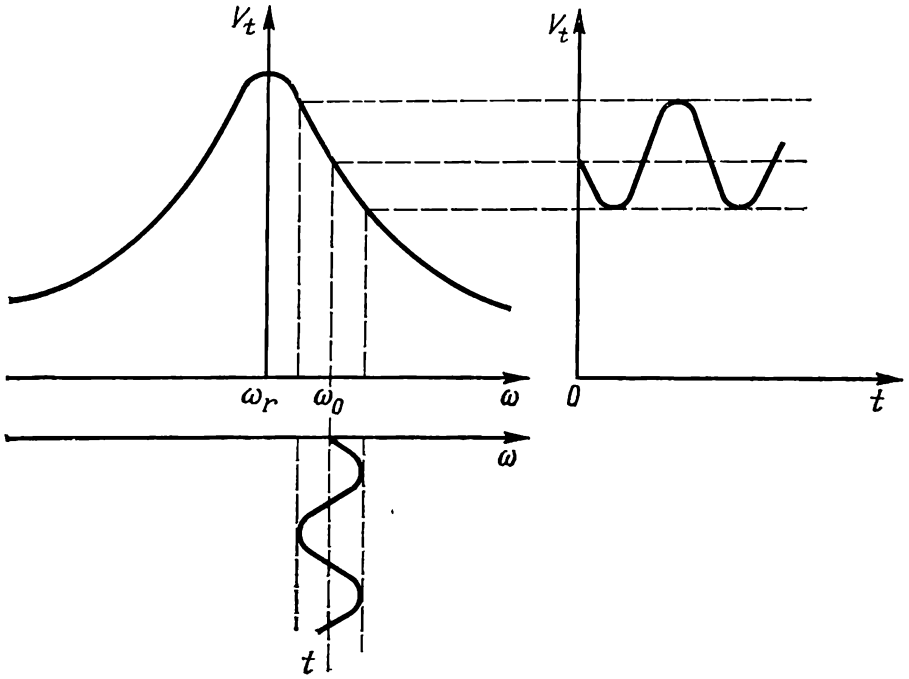


Fig. 8.37. Explaining the operation of the detector shown in Fig. 8.36

Figure 8.38 shows a schematic diagram of a frequency detector widely used in FM receivers and in devices for automatic frequency control in oscillators. It comprises an oscillatory circuit in the form

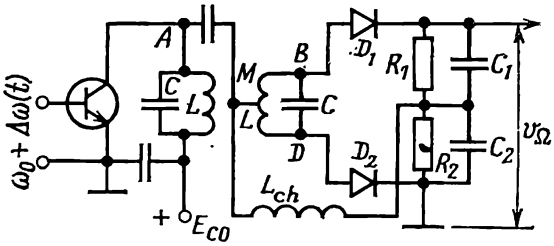


Fig. 8.38. Double-tuned frequency detector

of two inductively coupled tank circuits tuned to the frequency ω_0. The r-f voltage of frequency $\omega_0 \pm \Delta\omega$ is applied to the base of the transistor, and the demodulated voltage v_Ω is developed across the resistors R_1 and R_2. The inductor L_{ch} (choke) does not let pass through the r-f current. The principle of operation of this detector is

illustrated by the equivalent circuit and the vector diagram shown in Figs. 8.39 and 8.40.

Let $\mathbf{V}_1$ be the voltage across the first tank circuit and $\mathbf{V}_2$, that across the second tank circuit, and $\mathbf{V}_3$ and $\mathbf{V}_4$, the voltages at the points B and D with respect to the emitter (earth). It should be noted that $\mathbf{V}_3$ and $\mathbf{V}_4$ are the amplitudes of the r-f voltages applied to the diodes D_1 and D_2, respectively. In the absence of modulation, when the

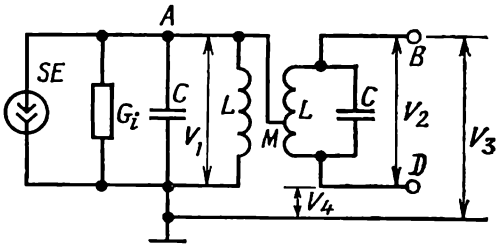


Fig. 8.39. Equivalent circuit of the selective network of a frequency detector (to Fig. 8.38)

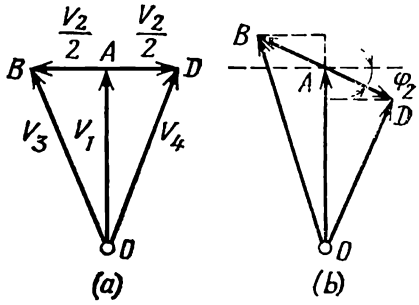


Fig. 8.40. Voltage vector diagram (to Fig. 8.39)

frequency of the input voltage coincides with the resonance frequencies of the tank circuits, the phase of the voltage $\mathbf{V}_2$ developed across the inductance of the second tank circuit is shifted 90° with respect to the resonance-frequency voltage $\mathbf{V}_{1r}$.

In fact, with inductive coupling of two identical tank circuits, we have

$$\mathbf{I}_2 = \frac{i(M/L)\,\mathbf{V}_{1r}}{Z_2}; \quad \mathbf{V}_2 = \mathbf{I}_2 i\omega L = \frac{M}{L}\frac{i\omega L}{Z_2}\mathbf{V}_{1r} \tag{8.61}$$

When $\omega = \omega_0 = \omega_r$, $Z_2 = r_2$ and $\omega_0 L/r_2 = Q_2$, so we have

$$\mathbf{V}_2 = i\,(M/L)\,Q_2\mathbf{V}_{1r} \tag{8.62}$$

i.e., the voltage $\mathbf{V}_2$ leads the voltage $\mathbf{V}_1$ by 90°.

Let us find the voltages $\mathbf{V}_3$ and $\mathbf{V}_4$. Taking into ccount that in the equivalent circuit the centre tap of the second tank circuit is connected directly to the point A and consequently, the voltage $\mathbf{V}_3$ is the sum of the voltage $\mathbf{V}_{1r}$ and half the voltage $\mathbf{V}_2$, we obtain

$$\mathbf{V}_3 = \mathbf{V}_{1r} + \frac{\mathbf{V}_2}{2} = \mathbf{V}_{1r} + i\frac{M}{L}Q_2\frac{\mathbf{V}_{1r}}{2} = \mathbf{V}_{1r}\left(1 + i\frac{M}{L}\frac{Q_2}{2}\right) \tag{8.63}$$

In a similar way, for the voltage $\mathbf{V}_4$ we may write

$$\mathbf{V}_4 = \mathbf{V}_{1r}\left(1 - i\frac{M}{L}\frac{Q_2}{2}\right) \tag{8.64}$$

The magnitudes of the voltages $\mathbf{V}_3$ and $\mathbf{V}_4$ are the same and are given by

$$V_3 = V_4 = V_{1r}\sqrt{1 + \left(\frac{M}{L}\frac{Q_2}{2}\right)^2} \tag{8.65}$$

and their phases are symmetrical with respect to the phase of the voltage $\mathbf{V}_{1r}$. The vector diagram corresponding to this case is shown in Fig. 8.40*a*. Since the rectified voltages V_{01} and V_{02} existing across the resistors R_1 and R_2 are proportional to the amplitudes V_3 and V_4, the resultant voltage at the detector output, equal to the difference between V_{01} and V_{02}, is zero at the resonance frequency.

Now let us discuss the voltage vector diagram in the case of detuning. Assume that the frequency at the detector input differs from the resonance frequency ω_0 by an amount $\Delta\omega$ so that $\Delta\omega/\omega_0 \ll 1$. Then the vector DB corresponding to the voltage $\mathbf{V}_2$ (Fig. 8.40*b*) will turn relative to its resonance position through an angle φ_2 which is defined by the expression

$$\varphi_2 = \arctan\left[(2\Delta\omega/\omega_0)\,Q_2\right] = \arctan a_2 \tag{8.66}$$

Instead of expressions (8.63) and (8.64) we obtain

$$\mathbf{V}_3 = \mathbf{V}_1\left[1 + \frac{i\,(M/2L)\,Q_2}{1+ia_2}\right] \tag{8.63'}$$

$$\mathbf{V}_4 = \mathbf{V}_1\left[1 - \frac{i\,(M/2L)\,Q_2}{1+ia_2}\right] \tag{8.64'}$$

The first and second tank circuits are usually made identical. Therefore, the ratio $M/L = k$ is the coupling coefficient of the circuits. In addition, we assume that $Q_1 = Q_2 = Q$ and $a_1 = a_2 = a$.

Introducing the designation

$$\beta = (M/L)\,Q = kQ$$

and changing to moduli, we get

$$V_3 = \frac{V_1}{2}\frac{\sqrt{4+(\beta+2a)^2}}{\sqrt{1+a^2}}, \qquad V_4 = \frac{V_1}{2}\frac{\sqrt{4+(\beta-2a)^2}}{\sqrt{1+a^2}}$$

Taking into account the differential connection of the loads and also formula (8.50), the output voltage is given by

$$v_\Omega = V_{01} - V_{02} = V_3\cos\theta - V_4\cos\theta$$

$$= V_1\cos\theta\,\frac{\sqrt{4+(\beta+2a)^2} - \sqrt{4+(\beta-2a)^2}}{2\sqrt{1+a^2}} \tag{8.67}$$

Taking into account the relation $V_1 \approx SEZ_{eq1}$, where Z_{eq1} is the equivalent impedance of the first tank circuit, defined by the

expression

$$Z_{eq1} = Z_{1r} \frac{(1+a^2)}{\sqrt{(1+a^2+\beta^2)^2+a^2(1+a^2-\beta^2)^2}}$$

$$= Z_{1r} \frac{(1+a^2)}{\sqrt{(1+\beta^2-a^2)^2+4a^2}}$$

formula (8.67) assumes the following final form:

$$v_\Omega \approx SZ_{1r} \cos\theta E \frac{\sqrt{4+(\beta+2a)^2}-\sqrt{4+(\beta-2a)^2}}{2\sqrt{(1+\beta^2-a^2)+4a^2}} = N\psi(a) \qquad (8.68)$$

where $N = SZ_{1r} \cos\theta E$ is a constant, while the function

$$\psi(a) = \frac{\sqrt{4+(\beta+2a)^2}-\sqrt{4+(\beta-2a)^2}}{2\sqrt{(1+\beta^2-a^2)+4a^2}} \qquad (8.69)$$

The function $\psi(a)$ is shown in Fig. 8.41 in the form of a family of characteristics for various values of the parameter β. Multiplying

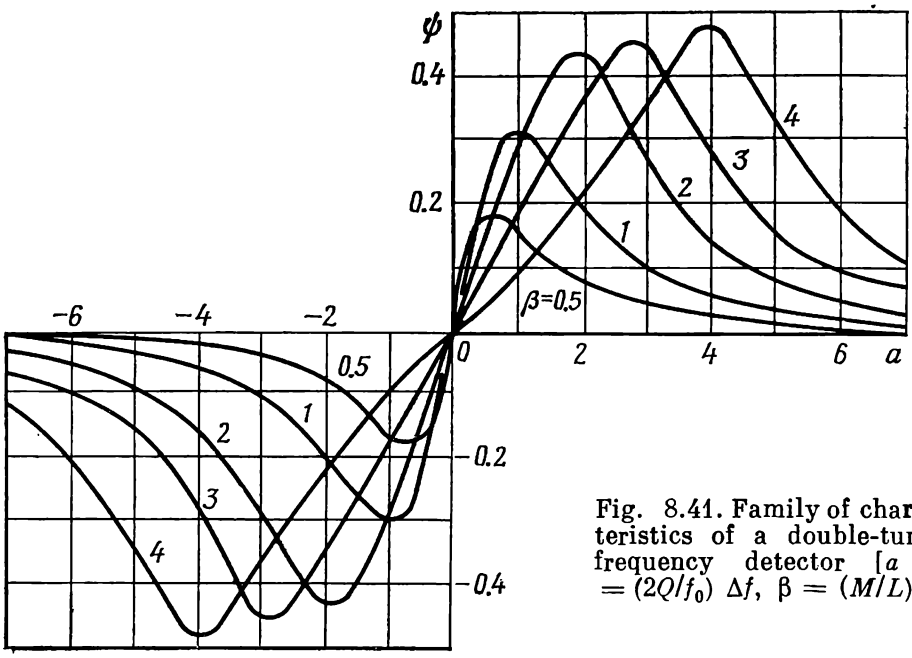


Fig. 8.41. Family of characteristics of a double-tuned frequency detector [$a = (2Q/f_0)\,\Delta f$, $\beta = (M/L)Q$]

the ordinates of these characteristics by N and the abscissae, by $f_0/2Q$, we obtain the characteristic of the frequency detector in the form of the relation between v_Ω [V] and Δf [Hz].

The main objective to be pursued when selecting the parameters of the tank circuits and the amount of coupling is to ensure that the characteristic of the frequency detector is linear and has the maximum possible slope.

Besides the above-discussed frequency detector circuits, there are a number of other systems differing only in details.

Let us now consider the principle of operation of a phase detector. Let the phase of an r-f oscillation to be detected change according to $\theta(t)$. If such an oscillation is applied to an ordinary frequency detector responding to the variation of the instantaneous frequency of the oscillation, the output voltage of the detector will be

$$v_{out}(t) = S_{fd}\,\Delta\omega(t) = S_{fd}\frac{d\theta(t)}{dt}$$

i.e., the output voltage will be proportional to the derivative of the phase of the input oscillation. From this it is clear that an ordinary frequency detector can be used for phase detection. It must only be supplemented with a correction circuit for integrating the output voltage, i.e., with a circuit having a frequency characteristic $\mathbf{K}(i\omega) = 1/i\omega\tau_0$. The simplest integrating devices were described in Sec. 6.5. Such a technique is used for detection of oscillations with a slowly varying phase, i.e., when the derivative of the phase is finite (e.g., when transmitting a speech). In the case of a stepwise phase variation and also, where it is necessary to compare the phase of the oscillation being received with that of a reference oscillation, special phase detectors are used, in which the output voltage is proportional to the envelope of the voltage obtained by adding together the oscillations whose phases are to be compared. Such devices are studied in special courses.

8.10. CONVERSION OF THE FREQUENCY OF A SIGNAL

In radio engineering, it is often required to shift the spectrum of a signal along the frequency axis by a definite constant amount, while preserving the structure of the signal. Such a shift is known as *frequency conversion*.

To reveal the main features of the process of frequency conversion, let us return to the problem of two voltages acting on a nonlinear element, which was briefly considered in Sec. 8.3. In this case, however, only one oscillation, namely that produced by the local oscillator, will be regarded as being harmonic. By the second oscillation we shall imply the signal to be converted, and it may be any complex but narrow-band process.

Thus, the nonlinear element is acted upon by two voltages:

— the voltage from the local oscillator

$$e_{lo} = E_{lo}\cos(\omega_{lo}t + \theta_{lo}) \tag{8.70}$$

— the voltage from the signal source

$$e_s = E_s(t)\cos\left[\int \omega_s(t)\,dt + \theta_s\right] \tag{8.71}$$

The amplitude E_{lo}, frequency ω_{lo} and epoch angle θ_{lo} of the local oscillator voltage are constant values, while the amplitude $E_s(t)$ and instantaneous frequency $\omega_s(t)$ of the signal may be modulated, i.e., they may be slow-varying functions of time (narrow-band process). The epoch angle θ_s of the signal is constant.

As in Sec. 8.8, the nonlinear element is a diode; however, its characteristic shall be approximated by a fourth-degree polynomial (and not by a second-degree polynomial as in Sec. 8.3) to completely determine the products of interaction of the signal and the local oscillator voltage:

$$\begin{aligned} i = i_0 + a_1(e_s + e_{lo}) + a_2(e_s + e_{lo})^2 + a_3(e_s + e_{lo})^3 \\ + a_4(e_s + e_{lo})^4 = i_0 + a_1 e_s + a_1 e_{lo} + a_2 e_s^2 \\ + \boxed{2a_2 e_s e_{lo}} + a_2 e_{lo}^2 + a_3 e_s^3 + \boxed{3a_3 e_s^2 e_{lo}} + \boxed{3a_3 e_s e_{lo}^2} + a_3 e_{lo}^3 \\ + a_4 e_s^4 + \boxed{6a_4 e_s^2 e_{lo}^2} + \boxed{4a_4 e_s e_{lo}^3} + \boxed{4a_4 e_s^3 e_{lo}} + a_4 e_{lo}^4 \end{aligned} \tag{8.72}$$

The terms containing only e_s or only e_{lo} are of no interest to us. From the point of view of frequency conversion (shift), only the terms containing the products of the form $e_s^n e_{lo}^m$ [enclosed in boxes on the right-hand side of expression (8.72)] are important. Substituting (8.70) and (8.71) into these products and rejecting all the components whose frequencies are other than the sum $\omega_s + \omega_{lo}$ or the difference $\omega_s - \omega_{lo}$, after some simple trigonometric manipulations, we get the following final result:

$$\begin{aligned} i_{\omega_s \pm \omega_{lo}}(t) = a_2 E_s(t) E_{lo} \left\{ \cos\left[\left(\int \omega_s(t)\, dt + \omega_{lo} t \right) + \theta_s + \theta_{lo} \right] \right. \\ \left. + \cos\left[\left(\int \omega_s(t)\, dt - \omega_{lo} t \right) + \theta_s - \theta_{lo} \right] \right\} \\ + \frac{3}{2} a_4 E_s(t) E_{lo} [E_s^2(t) + E_{lo}^2] \left\{ \cos\left[\left(\int \omega_s(t)\, dt + \omega_{lo} t \right) \right. \right. \\ \left. \left. + \theta_s + \theta_{lo} \right] + \cos\left[\left(\int \omega_s(t)\, dt - \omega_{lo} t \right) + \theta_s - \theta_{lo} \right] \right\} \end{aligned} \tag{8.73}$$

From this it is clear that the frequencies of interest, i.e., $\omega_s \pm \omega_{lo}$, appear only by virtue of the even powers of the polynomial approximating the characteristic of the nonlinear element. But only the quadratic term of the polynomial (that with the coefficient a_2) forms components whose amplitudes are *proportional to only the first power of* $E_s(t)$. The higher even powers (fourth, sixth, etc.) disturb this proportionality since the amplitudes of the oscillations caused thereby also contain powers of $E_s(t)$ that are higher than the first.

From this it follows that the amplitudes E_s and E_{lo} must be selected so that in expansion (8.72) the terms whose power does not

exceed the second can predominate. For this, the following inequalities must be satisfied:

$$E_s^2 \ll a_2 \Big/ \left(\frac{3}{2} a_4\right), \quad E_{lo}^2 \ll a_2 \Big/ \left(\frac{3}{2} a_4\right)$$

Then expression (8.73) transforms into the following one:

$$i_{\omega_s \pm \omega_{lo}}(t) \approx a_2 E_s(t)\, E_{lo} \left\{ \cos\left[\left(\int \omega_s(t)\, dt + \omega_{lo} t\right) + \theta_s + \theta_{lo}\right] + \cos\left[\left(\int \omega_s(t)\, dt - \omega_{lo} t\right) + \theta_s - \theta_{lo}\right]\right\} \qquad (8.74)$$

In radio receivers and other devices where frequency conversion is closely associated with amplification of signals, as a rule, $E_s \ll E_{lo}$.

The first term in the braces, whose frequency is $\omega_s(t) + \omega_{lo}$ (the derivative of the cosine argument), corresponds to the shift of

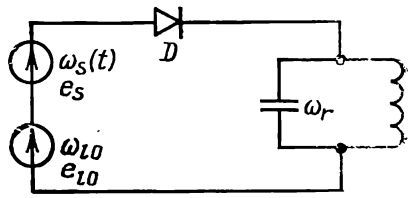


Fig. 8.42. Equivalent circuit of a frequency converter

the signal spectrum into the h-f region, while the second term with a frequency $\omega_s(t) - \omega_{lo}$ corresponds to the shift of the spectrum into the l-f region. Any of these frequencies (sum or difference) can be separated by using a suitable load at the converter output. For example, assume that the frequencies ω_s and ω_{lo} are very close and that it is required to separate the low (difference) frequency that is close to zero. This is a typical problem often met with in measuring systems (zero-beat method). In this case, the load must be the same as in amplitude detection, i.e., R and C must be connected in parallel so as to reject (filter out) the high frequencies ω_s and ω_{lo} and separate the difference frequency $|\omega_s - \omega_{lo}|$. If the difference frequncy $|\omega_s - \omega_{lo}|$ is within the r-f band, it should be separated by means of a resonance circuit (Fig. 8.42). For separation of the sum frequency $\omega_s + \omega_{lo}$, the tank circuit must be tuned to the frequency $\omega_r = \omega_s + \omega_{lo}$.

As a rule, the passband of the oscillatory circuit loading the converter is designed to handle the entire spectrum width of the modulated oscillation. In this case, all the current components whose frequencies are close to $|\omega_s \pm \omega_{lo}|$ are evenly transmitted through the circuit, and the structure of the output signal is identical with that of the input signal. The only difference consists in that the output frequency is equal to $\omega_s(t) + \omega_{lo}$ or $\omega_s(t) - \omega_{lo}$, depending on the resonance frequency of the load circuit.

Thus, in frequency conversion, the variations of the amplitude $E_s(t)$, frequncy $\omega_s(t)$, and phase $\int \omega_s(t)\,dt$ of the output oscillation are similar to the variations of those of the input oscillation. In this respect, the conversion of the signal is a *linear* process and the device is a linear converter or "mixer".

In conclusion, it should be noted that when the difference frequency is used, the structure of the signal is preserved completely only

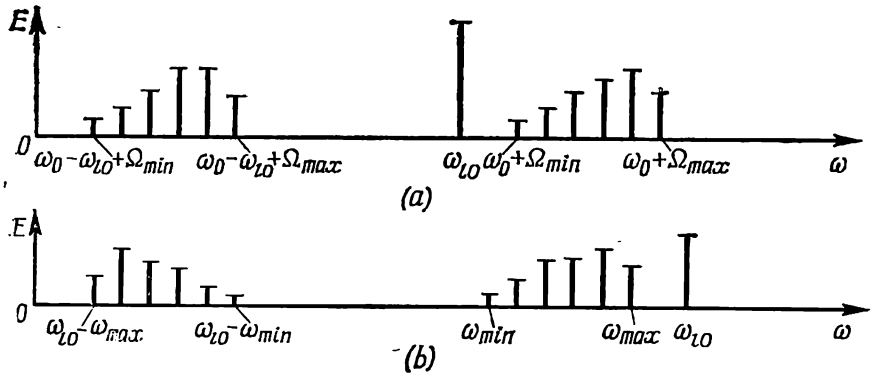


Fig. 8.43. Signal spectrum at the input and output of a converter (*a*) for $\omega_s > \omega_{lo}$; (*b*) for $\omega_s < \omega_{lo}$

where $\omega_s(t) > \omega_{lo}$. Where $\omega_s(t) < \omega_{lo}$, the signal spectrum is inverted.

Figure 8.43*a* shows a spectral diagram of a signal at the input and the output of a converter for the case where all the frequencies in the spectrum of the input oscillation are higher than the local-oscillator frequency ω_{lo}. The converted spectrum, that is the one shifted to the left by an amount ω_{lo}, has the same structure as the original spectrum. The case $\omega_{lo} > \omega_s(t)$ is shown in Fig. 8.43*b*. In the converted spectrum, ω_{max} and ω_{min} change places.

When converting the frequency of an ordinary amplitude-modulated oscillation consisting of two sidebands that are symmetrical relative to the centre frequency ω_0, there are no outward signs of spectrum inversion: simply the upper and lower sidebands change places. But when converting a frequency-modulated oscillation whose instantaneuos frequency is $\omega_s(t) = \omega_0 + \Delta\omega(t)$, the case $\omega_{lo} > \omega_s(t)$ results in that the instantaneous frequency of the output signal changes according to

$$|\omega_s(t) - \omega_{lo}| = \omega_{lo} - \omega_0 - \Delta\omega(t)$$

i.e., the sign before the frequency deviation $\Delta\omega(t)$ is changed.

From the above examples it is clear that the spectrum inversion in frequency conversion is to be taken into consideration only when the signal spectrum is asymmetrical with respect to the centre

frequency (in the case of frequency modulation, the asymmetry consists in that the signs before the side frequencies $\omega_0 - n\Omega$ are negative for odd n, see Sec. 4.6).

When converting the frequency of a signal with an asymmetrical spectrum, to preserve the structure of the signal, the local-oscillator frequency must be lower than the signal frequencies.

8.11. SYNCHRONOUS DETECTION

Let us consider a special type of conversion which takes place when the local-oscillator frequency is equal to the signal frequency. Setting in expression (8.74) $\omega_s = \omega_{lo}$ and restricting ourselves first to the unmodulated input oscillation ($E_s = E_0$), we obtain

$$i(t) = a_2 E_0 E_{lo} [\cos(2\omega_s t + \theta_s + \theta_{lo}) + \cos(\theta_s - \theta_{lo})] \qquad (8.75)$$

We see that in the particular case $\omega_{lo} = \omega_s$ the oscillation of the lower combination frequency degenerates into direct current

$$i_0 = a_2 E_{lo} \cos(\theta_s - \theta_{lo}) E_0 \qquad (8.76)$$

When $\theta_s - \theta_{lo} = 0$ or π, the current $| i_0 |$ is at its maximum and when $\theta_s - \theta_{lo} = \pi/2$, the current $i_0 = 0$.

By using a low-pass filter at the converter output, the frequency $2\omega_s$ is rejected and only a d-c voltage proportional to the current i_0 remains at the filter output. In the presence of amplitude modulation, when $e_s(t) = E_s(t)\cos(\omega_s t + \theta_s)$, the output oscillation will be proportional to the current

$$i_\Omega(t) = a_2 E_{lo} \cos(\theta_s - \theta_{lo}) E_s(t) \qquad (8.77)$$

i.e., it will follow the modulating function of the r-f oscillation $e_s(t)$.

In other words, the transmitted message is produced at the converter output, the signal processing being essentially *linear* with respect to the input oscillation, provided that $E_s(t) \ll E_{lo}$.

The main advantage of such a method of processing, proposed in 1934 by E. G. Momot and called *synchronous detection*, is the high selectivity of reception of weak radio signals on a noisy background (the interaction of the signal with noise in a nonlinear device, such as an ordinary amplitude detector, is eliminated).

It should be noted, however, that the realization of the principle of synchronous detection is associated with considerable difficulties, for the synchronization of the local-oscillator frequency with the frequency of the signal being received is a complex problem, especially when receiving weak signals on a noisy background.

8.12. GENERATION OF AMPLITUDE-MODULATED OSCILLATIONS

In Sec. 8.4 it was indicated that when two harmonic oscillations of frequencies ω_1 and ω_2 satisfying the condition $\omega_1 \ll \omega_2$ act upon a nonlinear resistive element with a square-law characteristic, the

current spectrum contains, among other spectrum components, three frequencies — ω_2, $\omega_2 + \omega_1$, and $\omega_2 - \omega_1$ — that form the spectrum of an amplitude-modulated oscillation.

An important requirement usually imposed on oscillators and transmitters is for them to have high power output along with good efficiency. It is clear that the square-law operation of a nonlinear element does not meet this requirement. To improve the modulation efficiency, the resistive nonlinear element must in fact operate

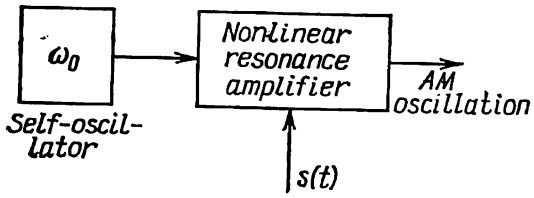


Fig. 8.44. Block diagram of a device for producing AM oscillations

under essentially nonlinear conditions, i.e. with current cutoff. Therefore, the amplitude modulation of an r-f oscillation is reduced to the action upon a *nonlinear resonance amplifier*. A block diagram of a system for producing amplitude-modulated oscillations is shown in Fig. 8.44.

The carrier oscillation of frequency ω_0 from an independent source (self-oscillator) is applied to the input of the nonlinear resonance

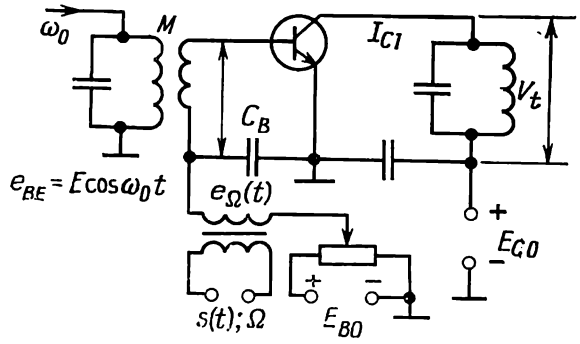


Fig. 8.45. Circuit diagram to Fig. 8.44

amplifier. The modulating oscillation (message) $s(t)$, whose spectrum is located in the range of frequencies that are low compared to ω_0, changes the position of the operating point on the current-voltage characteristic curve of the nonlinear element, thus changing the amplitude at the output.

One of the possible circuits for applying a modulating oscillation $s(t)$ to a resonance (transistor) amplifier is shown in Fig. 8.45. The

capacitor C_B in the base-to-emitter circuit protects the l-f circuit from r-f currents.

The operation of the nonlinear amplifier in the case of tone modulation [$s(t)$ is a harmonic function with a frequency Ω] is illustrated in Fig. 8.46*a*.

Since the collector current $i_C = \beta i_B$, the amplitude of the voltage across the tank circuit, produced by the first harmonic of the collector current, is

$$V_t = I_{C1} Z_{eq,\ r} = \beta I_B Z_{eq,\ r}$$

The plot in Fig. 8.46 differs from the one discussed in Sec. 8.3 (Fig. 8.10) in that the amplitude I_m of the current pulses (Fig. 8.46*b*)

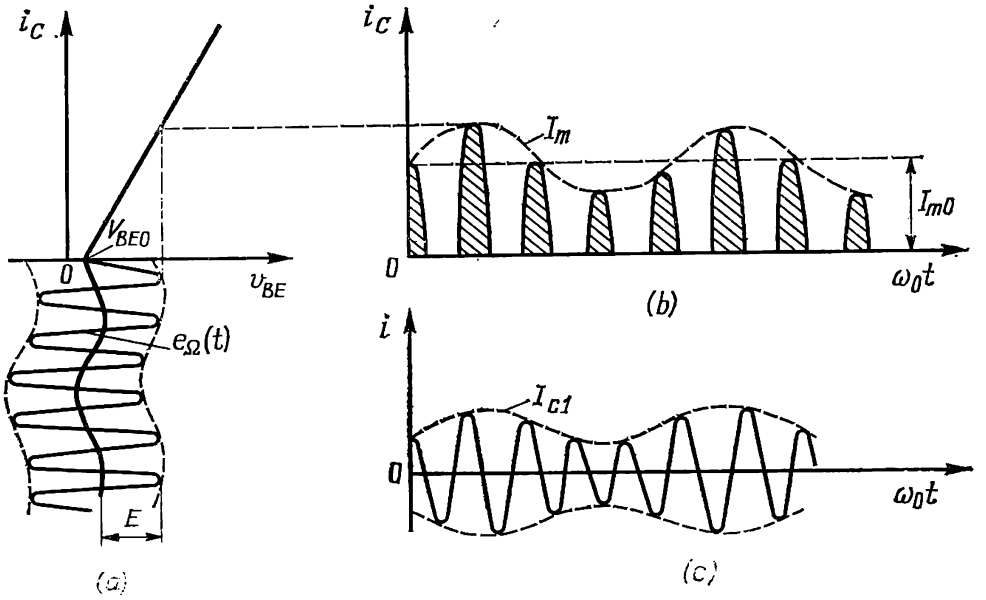


Fig. 8.46. Operation of a nonlinear resonance amplifier in amplitude modulation

depends on the modulating voltage $e_\Omega(t)$. This results in a change in the amplitude of the first harmonic of the collector current and consequently, in a change in the amplitude of the voltage across the tank circuit of the amplifier. The amplitude-modulated current of the centre frequency ω_0 is shown in Fig. 8.46*c*. The dashed line shows the variation of I_{C1} — the amplitude of the first harmonic of the collector current. With the amplitude of the modulating voltage being correctly chosen, the change ΔI_m in the amplitude of the current pulses with respect to the initial value I_{m0} is related to e_Ω by the linear relation

$$\Delta I_m = k_{am} e_\Omega$$

where k_{am} is a constant coefficient.

Since the change in e_Ω (with the amplitude E of the r-f oscillation being constant) is accompanied by a change in the cutoff angle θ, the amplitude of the first harmonic of the collector current [see formulas (8.24) and (8.26)] is given by

$$I_{C1} = \alpha_1(\theta)\,(I_{m0} + \Delta I_m) = I_{10} + k_{am}\alpha_1(\theta)\,e_\Omega$$

where I_{10} is the amplitude of the carrier oscillation. The product $k_{am}\,\alpha_1\,(\theta)\,e_\Omega$ varies according to a law *differing* from the law of variation of the modulating voltage. From this it is clear that bias modulation is inevitably accompanied by distortion: the function describing I_{C1} differs in form from that describing the voltage e_Ω. The distortion can be reasonably small, provided the cutoff angle is chosen correctly and modulation is shallow [(40 to 50)%].

Chapter 9

HARMONIC SELF-OSCILLATORS

9.1. SELF-OSCILLATING SYSTEM

Any self-oscillator is a nonlinear device converting the supply energy into the energy of oscillations. Regardless of its type and field of application, any self-oscillator must have a power-supply source, an amplifier, and a feedback circuit. From the information given in Ch. 5 it follows that feedback here must be positive.

This chapter deals mainly with the study of the phenomena taking place in self-oscillators used for producing high-frequency harmonic oscillations. Such oscillators use electron tubes, transistors, and other similar devices for amplifying elements and oscillatory circuits with lumped or distributed parameters for load circuits.

A self-oscillator operating under steady-state conditions is an ordinary nonlinear amplifier excited by the oscillations generated in the oscillator itself, the oscillations from the output of the amplifier being fed to its input via the feedback circuit. If the oscillation amplitude and phase satisfy certain definite conditions, the behaviour of a self-oscillator from the energy standpoint is identical with that of an externally excited oscillator. However, the self-oscillator has essential specific features. The frequency and amplitude of a *self-oscillation* under steady-state conditions are determined solely by the parameters of the oscillator itself, while in the other case, the frequency and amplitude are set forcibly by the exciter. What is more, in the case of self-excitation, the mechanism of development of oscillations when starting the self-oscillator is of great importance. These features of the self-oscillator can be revealed by examining its behaviour during the entire rise time of oscillations from the start-up and till steady state is reached. The following picture may be outlined. At the moment the self-oscillator is started, in its oscillatory circuit there develop free oscillations due to switching surges, electric fluctuations, etc. Owing to positive feedback, these initial oscillations are amplified, the amplification at the initial stage, while the oscillation amplitude is still low, being practically linear, so that the circuit may be regarded as linear. From the energy standpoint, the process of rise of the oscillation amplitude is explained by the fact that the amount of energy supplied by the amplifier to the load is greater than the amount of energy dissipated in the load during the same period of time. As the oscillation amplitude grows

higher, the nonlinearity of the device (the curvature of the current-voltage characteristic of the amplifying element) starts manifesting itself, and amplification decreases. The oscillation amplitude stops rising once the decreasing amplification has reached the level at which the damping of the oscillations in the load is just compensated for, the energy supplied by the amplifier during one oscillation period being equal to the energy dissipated in the load during the same time.

Thus, at the last stage of the process of setting in of oscillations, to the fore comes the nonlinearity of the circuit, and without taking

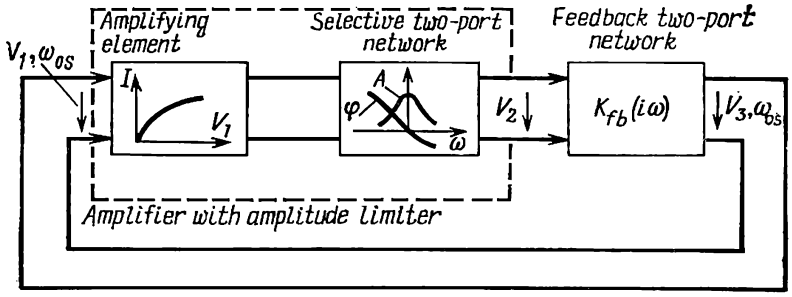


Fig. 9.1. Block diagram of a self-oscillator

account of this factor one cannot determine the steady-state parameters of the self-oscillator.

Any radio-frequency self-oscillator operating under steady-state conditions can be represented in the form of the block diagram shown in Fig. 9.1, where ω_{os} stands for the oscillation frequency. The oscillator consists of three two-port networks — a lag-free nonlinear network and two linear networks. The nonlinear two-port network corresponds to the amplifying element (an electron tube, transistor, klystron, etc.), the first linear network corresponds to the oscillatory circuit, and the second network, to the feedback circuit of the oscillator. Such a representation is valid for systems with *external feedback*.

In Sec. 9.8 we shall discuss some oscillators whose operation depends on *internal feedback* which calls for a somewhat different treatment of the generalized circuit.

The amplifying element in conjunction with the selective two-port network that filters out (suppresses) the higher harmonics is a conventional nonlinear amplifier generating a sinusoidal voltage at its output. In the general case, amplification depends both on the frequency ω_{os} (due to the selectivity of the two-port network) and on the amplitude V_1 (due to the nonlinearity of the amplifying element). Let us denote the amplification factor of this device as

$\mathbf{K}_a\,(i\omega_{os}, V_1)$. It is clear that

$$\mathbf{K}_a\,(i\omega_{os},\ V_1) = \mathbf{V}_2/\mathbf{V}_1 \tag{9.1}$$

For a fixed frequency ω_{os}, the amplification factor $\mathbf{K}_a$ is a function of only the amplitude V_1.

The transmission coefficient of the linear feedback two-port network, which will further be referred to as the feedback factor, can be expressed through the amplitudes $\mathbf{V}_3$ and $\mathbf{V}_2$ as follows:

$$\mathbf{K}_{fb}\,(i\omega) = \mathbf{V}_3/\mathbf{V}_2$$

But the voltage $\mathbf{V}_3$ at the output of the feedback two-port network is at the same time the input voltage $\mathbf{V}_1$ of the amplifier. Consequently

$$\mathbf{K}_{fb}\,(i\omega) = \mathbf{V}_1/\mathbf{V}_2 \tag{9.2}$$

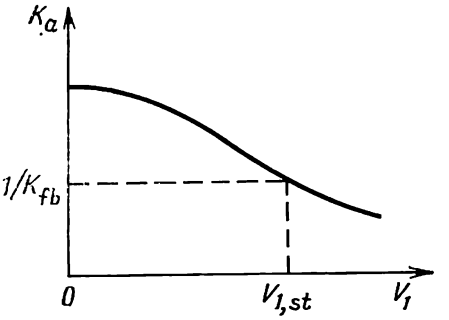


Fig. 9.2. Determining the steady-state oscillation amplitude

Comparing this expression with relation (9.1), we come to a conclusion that when the oscillator is under steady-state conditions (i.e., when the method of complex amplitudes can be used), the factors $\mathbf{K}_a\,(i\omega_{os}, V_1)$ and $\mathbf{K}_{fb}\,(i\omega_{os})$ are mutually reciprocal:

$$\mathbf{K}_a\,(i\omega_{os},\ V_1)\,\mathbf{K}_{fb}\,(i\omega_{os}) = 1 \tag{9.3}$$

Since the transmission coefficient of a linear two-port network is independent of the oscillation amplitude, expression (9.3) can be used for determining the steady-state oscillation amplitude at a given K_{fb}. As mentioned above, the amplitude stops increasing exactly at the instant when K_a, while decreasing with an increase in the oscillation amplitude (due to the nonlinearity of the current-voltage characteristic of the amplifying element), becomes equal to $1/K_{fb}$. This is illustrated in Fig. 9.2. The steady-state oscillation amplitude V_{1st} is determined as the abscissa of the point of intersection of the curve K_a and the horizontal line drawn at a level of $1/K_{fb}$. On the other hand, expression (9.3) can be used for determining the feedback factor required for maintaining a certain definite oscillation amplitude V_{1st} for a given function $K_a(V_1)$.

While revealing these general properties of the self-oscillator, there has been no need in specifying either the type of the amplifying element or the circuit arrangement of the oscillator. This is explained by the fact that we have restricted ourselves to the consideration of the *steady-state* operation of the self-oscillator. But to reveal the mechanism of origination of oscillations and also that of

setting in of steady-state conditions, one must proceed from the actual electron device and the actual oscillator circuit.

Let us emphasize an important requirement placed on self-oscillators intended for use in information-transmitting devices, namely: the generated oscillations must be strictly monochromatic (in the absence of modulation). Any disturbance of monochromaticity manifesting itself in the form of spurious changes in the oscillation amplitude, frequency or phase may be the cause of noise in the radio communication channel. The requirement for monochromaticity also includes the requirement for the frequency stability of self-oscillation.

9.2. ORIGINATION OF OSCILLATIONS IN A SELF-OSCILLATOR

A vacuum-tube self-oscillator circuit (Fig. 9.3*a*) proves most convenient for the study of the mechanism of origination and rise of oscillations.

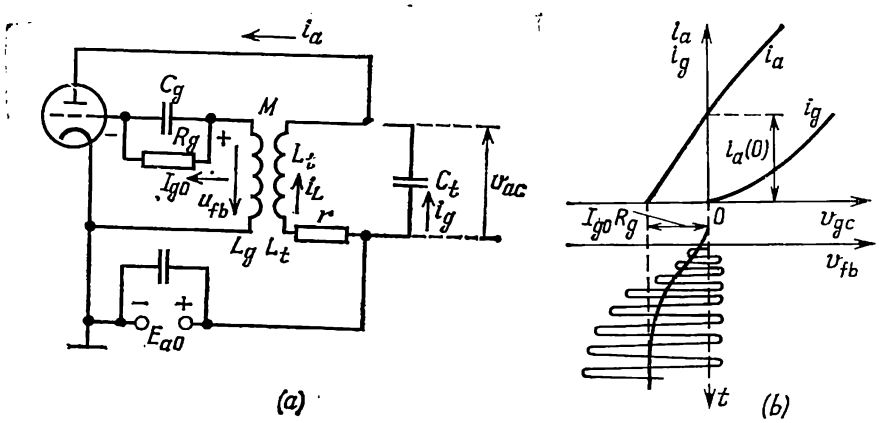


Fig. 9.3. (*a*) Single-tuned vacuum-tube self-oscillator and (*b*) its operation during starting

Assume that the self-oscillator is started by switching on the d-c voltage E_{a0} at the moment $t = 0$. The surge of the anode current i_a (0) (Fig. 9.3*b*) excites in the L_t-C_t-r tank circuit a free oscillation whose parameters are defined by the parameters of the tank circuit, tube, and feedback. At the initial stage of the start-up, while the oscillation amplitude is still low, the circuit shown in Fig. 9.3*a* can be regarded as linear. Let us set up for this circuit a differential equation taking into account only the a-c components of current and voltages.

The oscillatory voltage v_{ac} across the tank circuit and the currents i_L, i_C, i_a (Fig. 9.3*a*) are interrelated by the obvious relations

$$i_a = i_L + i_C, \quad i_C = C_t \frac{dv_{ac}}{dt}, \quad v_{ac} = ri_L + L_t \frac{di_L}{dt} \qquad (9.4)$$

Let us take, for example, the current i_L in the inductive branch of the circuit to be the function sought for. Eliminating i_C from the first equation in (9.4) and using the second and third equations, we obtain

$$i_a = i_L + rC_t \frac{di_L}{dt} + L_t C_t \frac{d^2 i_L}{dt^2} \tag{9.5}$$

Now it is necessary to express the current i_a through the voltages applied to the tube electrodes. Under linear conditions, we may use for this purpose an expression similar to (5.40):

$$i_a = Sv_{gc} - v_{ac}/R_i \tag{9.6}$$

In the system in question, the voltage v_{gc} is the feedback voltage, so $v_{gc} = v_{fb} = Mdi_L/dt$ (M is taken with the plus sign, because feedback is positive) hence, expression (9.6) may be written in the form

$$i_a = SM \frac{di_L}{dt} - \frac{v_{ac}}{R_i} \tag{9.7}$$

Equating the right-hand sides of equations (9.5) and (9.7) and taking into account the third equation in (9.4), we get, after collecting terms, the following differential equation:

$$\frac{d^2 i_L}{dt^2} + \left(\frac{r}{L_t} + \frac{1}{C_t R_i} - \frac{SM}{C_t L_t}\right) \frac{di_L}{dt} + \frac{1 + r/R_i}{L_t C_t} i_L = 0 \tag{9.8}$$

The quantity $(-SM/C_t)$ in the coefficient multiplying the first derivative has the meaning of the *negative resistance* introduced into the tank circuit by the amplifier with positive feedback.

Denoting

$$\alpha_{eq} = (1/2L_t)\,(r + L_t/C_i R_i - SM/C_t) \tag{9.9}$$

we reduce (9.8) to the form

$$\frac{d^2 i_L}{dt^2} + 2\alpha_{eq} \frac{di_L}{dt} + \frac{(1 + r/R_i)}{L_t C_t} i_L = 0 \tag{9.9'}$$

The general solution of equation (9.9') has the form

$$i_L(t) = A_0 e^{-\alpha_{eq} t} \cos(\omega_{fr} t + \theta_0) \tag{9.10}$$

where the amplitude A_0 and phase angle θ_0 are constants depending on the initial conditions, and the frequency of free oscillation

$$\omega_{fr} = \sqrt{(1 + r/R_i)/L_t C_l - \alpha_{eq}^2} \tag{9.11}$$

Solution (9.10) makes sense if the following inequality is satisfied:

$$(1 + r/R_i)/L_t C_t > \alpha_{eq}^2$$

The character of change of the free-oscillation amplitude (9.10) depends on the sign of α_{eq}, i.e., on the sign of the coefficient multi-

plying the first derivative in equation (9.8). If $\alpha_{eq} > 0$, oscillation is damped (Fig. 9.4*a*), if $\alpha_{eq} < 0$, the oscillation amplitude rises (Fig. 9.4*b*).

Taking into account expression (9.9), we arrive at the following condition for the origination and rise of the oscillation:

$$\alpha_{eq} = (1/2L_t)\,(r + L_t/C_tR_i - SM/C_t) < 0$$

or

$$SM/C_t > r + L_t/C_tR_i \qquad (9.12)$$

The satisfaction of these conditions provides for the rise of the oscillation amplitude, no matter how low its initial values.

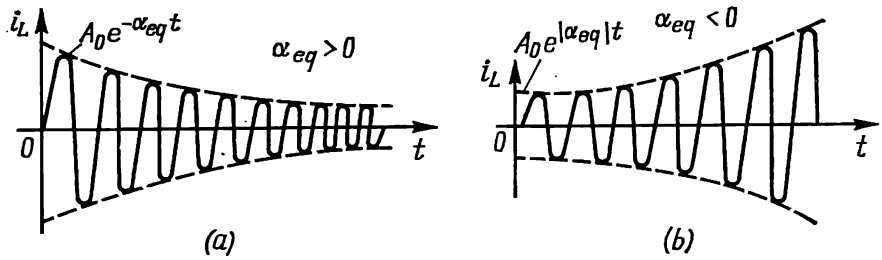


Fig. 9.4. Change of the free-oscillation amplitude depending on the sign of α_{eq}

Inequality (9.12) can be made more obvious by writing it in the form

$$M/L_t > rC_t/SL_t + 1/SR_i$$

Taking into account that the ratio M/L_t, equal to the ratio of the voltage v_{fb} to the voltage v_{ac}, is the feedback factor K_{fb} and also that $rC_t/L_t = 1/Z_{eq,\,r}$ and $1/SR_i = D$, we obtain

$$K_{fb} > 1/SZ_{eq,\,r} + D \qquad (9.13)$$

This inequality is the main *condition for self-excitation of an oscillator*. It enables one to explain easily the effect of the basic parameters of the amplifying device and the circuit arrangement of the self-oscillator on the process of origination of oscillations. The greater the slope S of the current-voltage characteristic, the lower the value of K_{fb} required and the easier the self-oscillation to start. On the contrary, an increase in the parameter D reflecting the reaction of the anode voltage on the input circuit requires that K_{fb} should be increased. It is also obvious that an increase in the resistance r accounting for the losses in the circuit and reducing the value of $Z_{eq,\,r} = L_t/rC_t$ requires that feedback be increased.

Note that the right-hand side of expression (9.13)

$$\frac{1}{SZ_{eq,\,r}} + D = \frac{1 + DSZ_{eq,\,r}}{SZ_{eq,\,r}} = \frac{R_i + Z_{eq,\,r}}{SR_iZ_{eq,\,r}}$$

is none other than the inverse of the amplification factor under linear conditions [see expression (5.42)].

Thus, inequality (9.13) may be written in another form:

$$K_{fb} > 1/K_a \qquad (9.13')$$

The same result can also be obtained by regarding a self-oscillator at the initial stage of the start-up as a linear amplifier with positive feedback. With $K_{fb}K_a > 1$, such an amplifier is *unstable* (see Sec. 5.10).

As the oscillation amplitude rises, the coefficient K_a decreases. This is due to the negative bias voltage $V_{gc0} = R_g I_{g0}$ developed across the resistor R_g by the d-c component of the grid current I_{g0} (Fig. 9.3*a*). The phenomena occurring in the grid circuit are exactly the same as those in a half-wave rectifier (see Sec. 8.7) whose diode is represented by the grid-to-cathode clearance of the tube and the load, by the R_g-C_g circuit. With the time constant $R_g C_g$ much in excess of the period of the r-f oscillation v_{fb} (t), the rectified voltage rises in proportion to the amplitude of the feedback voltage v_{fb} (t). As a result, as the oscillation amplitude increases, the operating point on the tube characteristic curve moves to the left, this leading to anode current cutoff and reduction of the slope S_{av} (see Sec. 8.4).

Steady-state self-oscillation conditions set in when inequality (9.13′) turns into equality.

Thus, the R_g-C_g circuit provides for automatic variation of the bias voltage, thus making it possible to combine favourable starting conditions ($V_{gc0} = 0$) with good efficiency of the oscillator (anode current cutoff) under steady-state conditions.

The inequality $K_{fb} > 1/K_a$ may be regarded as a condition for self-excitation of any self-oscillator. However, the mechanism of limiting the oscillation amplitude depends on specific features of the amplifying device. Thus, in a common-emitter transistor oscillator (Fig. 9.5*a*), at the moment of starting the operating point on the current-voltage characteristic curve is located not at the origin of coordinates, but at a point corresponding to a positive value of V_{BE0} (Fig. 9.5*b*). This is necessary because the collector and base currents in a transistor are related by the relation $i_C = \beta i_B$ and to meet the requirement for the great slope of the characteristic curve (to facilitate self-excitation conditions), the operating point must be set on the linear portion of the characteristic i_B (v_{BE}). Therefore, at the initial stage of starting the self-oscillator, the rise in the oscillation amplitude is not accompanied by an increase in the bias voltage (negative). The operating point is shifted to the left only when the oscillation amplitude comes to the lower bend of the curve where the effect of rectification of the voltage v_{fb} (t) in the base-to-emitter circuit starts manifesting itself.

In contrast to the circuit in Fig. 9.3*a*, that in Fig. 9.5*a* is provided with an independent source of the d-c voltage E_{B0}, that is connected in series into the R_B-C_B automatic bias-control circuit.

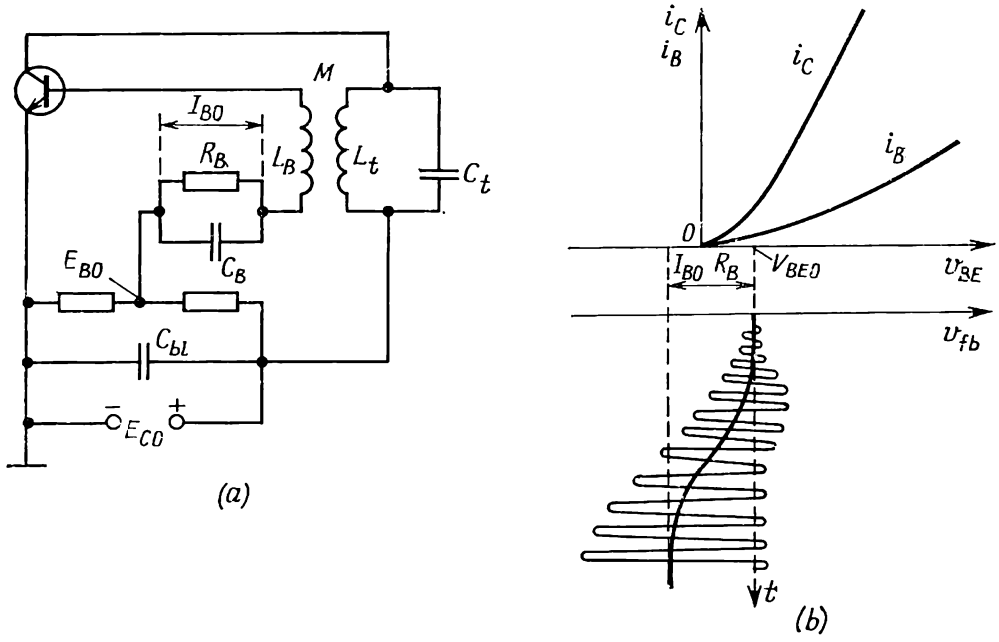


Fig. 9.5. (*a*) Single-tuned transistor self-oscillator and (*b*) its operation during starting

In transistor self-oscillators, the voltages E_{B0} and E_{C0} are usually supplied from a common source.

9.3. STEADY-STATE OPERATION OF A SELF-OSCILLATOR. PHASE BALANCE

Having revealed the conditions necessary for the origination of oscillations, let us determine the amplitude and frequency of the self-oscillations under steady-state conditions. The amplitude can be found using relation (9.3) applicable to any self-oscillator. Inequality (9.13) turns into equality (9.3) only when the average slope S_{av} has reduced so as to satisfy the condition

$$K_{fb} = 1/S_{av}Z_{eq,\,r} + D \quad \text{or} \quad S_{av} = 1/(K_{fb} - D)\,Z_{eq,\,r} \qquad (9.14)$$

Since S_{av} depends on the oscillation amplitude, the second equation in (9.14) makes it possible to find the steady-state amplitude. The determination of the steady-state amplitude, based on the method of *oscillatory characteristic* I_t (E_{in}) (where I_t is the current

amplitude in the tank circuit of the amplifier obtained from the oscillator by removing feedback), is more graphic.

By specifying the amplitude E_{in} of an r-f oscillation of frequency $\omega = \omega_r = 1/\sqrt{L_t C_t}$ at the input of the amplifier, the amplitude I_t of the current in the tank circuit is found theoretically or experimentally. A typical oscillatory characteristic is shown in Fig. 9.6 (curve *I*). For low amplitudes E_{in}, this characteristic is linear, since the operating point is located on the linear portion of the current-

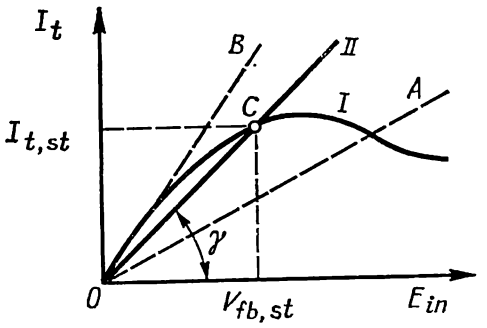


Fig. 9.6. Oscillatory characteristic of a nonlinear amplifier with automatic bias

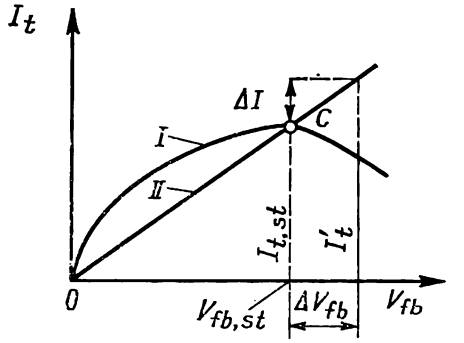


Fig. 9.7. Illustrating the stability of the steady-state operating conditions of a self-oscillator

voltage characteristic. The limitation of the oscillatory characteristic of the amplifier in the case of high amplitudes E_{in} is due to the rise of the bias voltage (if the automatic-bias circuit is used, see the preceding section).

To determine the current amplitude that would be established in the self-oscillator (after introducing feedback), it is necessary to find the relation between I_t and the voltage V_{fb}. Since $V_{fb} = I_t X_{coup}$, where X_{coup} is the coupling reactance, $I_t = V_{fb}/X_{coup}$. This relation is determined by the linear circuit of the oscillator and is shown in Fig. 9.6 in the form of the line *II* inclined to the abscissa axis at an angle $\gamma = \arctan(1/X_{coup})$. This line is called the *feedback line.*

The ordinate of the point of intersection of the lines *I* and *II* gives the steady-state current amplitude $I_{t,st}$, while the abscissa gives the steady-state voltage amplitude $V_{fb,st}$. In fact, at the point of intersection, the current I_t produced in the tank circuit by the amplifying device (line *I*) coincides with the current (line *II*) required for the development of the initial voltage V_{fb}.

As the coupling becomes stronger, the inclination of the line *II* decreases and the steady-state current amplitude increases. With a very strong feedback, the amplitude $I_{t,st}$ may even decrease due to the drop of the oscillatory characteristic of the amplifier, which

occurs when the excitation amplitude enters the saturation region of the current-voltage characteristic of the amplifying device. Such conditions are obtained when the coupling corresponds to the line OA in Fig. 9.6.

It can easily be shown that the point C of intersection of the lines I and II is *stable*. This means that in the case of random fluctuations of the current amplitude with respect to the steady-state value the self-oscillator returns to its initial condition. This property of the self-oscillator is illustrated in Fig. 9.7. Assume that the current amplitude in the tank circuit has increased by ΔI. This will cause an increase in the feedback voltage by ΔV_{fb}. However, with an input voltage of $V_{fb,st} + \Delta V_{fb}$, the amplifying device can circulate through the tank circuit only a current I_t' that is smaller than the current $I_{t,st} + \Delta I$. Therefore, the current in the tank circuit cannot be kept at a level of $I_{t,st} + \Delta I$ and must drop down, i.e., return to the initial value $I_{t,st}$. The same will take place in the case of a random decrease of the current in the tank circuit.

Let us find the frequency of the self-oscillations. As a first approximation, this frequency coincides with the natural frequency of the L_t-C_t-r tank circuit shunted by the internal resistance of the electron device. In linear approach (when considering the initial stage of the amplitude rise), the shunting effect has been taken account of by the coefficient $(1 + r/R_i)$ multiplying the last term in equation (9.8).

Under steady-state conditions, when the internal resistance R_i' of the electron device, referred to the current of the first harmonic, depends on the cutoff angle (see Sec. 8.4), the oscillation frequency is defined by the expression

$$\omega_{os} = (1/\sqrt{L_t C_t})\sqrt{1 + r/R_i'} \tag{9.15}$$

This frequency correction has to be taken into account when evaluating the frequency instability due to the unstable operation of the amplifying device. But in technical calculations, the self-oscillation frequency is usually taken to be equal to the resonance frequency of the tank circuit.

However, there are other factors that have a stronger effect on the oscillator frequency than R_i'. To reveal these factors, let us consider the phase relations in the closed feedback loop of the self-oscillator. The sum of all the phase shifts in the loop must be equal to $n2\pi$, where n is an integer [see (5.98)]. This condition defines the phase balance in the self-oscillator.

For a simple single-tuned self-oscillator, this condition may be written in the form

$$\varphi_a + \varphi_{fb} = 2\pi \tag{9.16}$$

where φ_a stands for the argument of the complex amplification factor $\mathbf{K}_a$ and φ_{fb}, for the argument of the complex feedback factor $\mathbf{K}_{fb}$.

Using the expression for the amplification factor

$$\mathbf{K}_a \approx -\mathbf{S}_{av} Z_{eq,\,r}(i\omega) \tag{9.17}$$

where $\mathbf{S}_{av}$ is generally the complex slope, we get for φ_a the following expression

$$\varphi_a = \varphi_S + \varphi_Z + \pi \tag{9.18}$$

Here, φ_S is the argument of $\mathbf{S}_{av}$ and φ_Z, the argument of the impedance of the parallel tank circuit. The term π takes into account the minus sign on the right-hand side of (9.17).

Thus, phase-balance equation (9.16) for the single-tuned self-oscillator assumes the form

$$\varphi_S + \varphi_Z + \varphi_{fb} + \pi = 2\pi$$

or

$$\varphi_S + \varphi_Z + \varphi_{fb} = \pi \tag{9.19}$$

From condition (9.19) it follows that all the factors that have influence on the phase shifts in the individual links of the self-oscillator also affect the oscillation frequency. For example, the insertion of a phase-shifting circuit into feedback two-port network displaces the oscillation frequency with respect to the resonance frequency of the tank circuit of the self-oscillator. The operation of such a self-oscillator, in which a delay line is used as a phase shifter, is described in detail in Sec. 9.9.

In practice, one often has to take into account the effect of the angle φ_S on the oscillation frequency. In all preceding sections of this chapter, as well as in Ch. 8, the average slope of an amplifying device was considered a real number ($\varphi_S = 0$). However, we should mention at least two factors which impart the average slope a complex character: (1) incomplete suppression of the higher harmonics of pulsed current and (2) inertia of electrons.

The mechanism of the effect of the higher-harmonic currents on the oscillation frequency consists in the following. These currents, while flowing through the tank circuit, produce a voltage drop, though very small, so that the resultant voltage across the tank circuit (and hence, at the output of the feedback network) becomes nonsinusoidal. As a result, the positive half-wave of the exciting voltage determining the shape of the current pulse deforms and becomes asymmetric with respect to its maximum value. The asymmetry is due to the fact that for the higher current harmonics the tank circuit is almost purely reactive, while for the first harmonic it is resistive, so the epoch angle of the additional voltages due to the higher harmonics is 90° when that of the first-harmonic voltage is zero.

The asymmetry of the current pulse, in its turn causes some shift of the phase of the first-harmonic current with respect to the first harmonic of the exciting voltage. As a result, the ratio of I_1 to E_1, i.e., the average slope S_{av}, becomes a complex number. It is clear that the higher the Q-factor of the tank circuit, the more harmonic the shape of the voltages and the weaker the effect of the higher harmonics on the oscillation frequency.

In oscillators having ordinary tank circuits, the relative frequency correction due to the effect of the higher harmonics is of the order of 10^{-4} to 10^{-5}.

The second of the above factors, i.e., the effect of the electron inertia, is significant only in self-oscillators operating at very high frequencies, where the transit time of the electron in the interelectrode spaces is commensurable with the oscillation period. This results in a substantial phase shift between the first current harmonic and the voltage at the input of the electron device. This shift should be taken into account when designing the feedback circuit.

9.4. SOFT AND HARD SELF-EXCITATION CONDITIONS

Now let us return to Fig. 9.6 and reveal the behaviour of the self-oscillator when varying the feedback factor. As the coupling is weakened, the inclination of the line II increases and at a certain critical value of K_{fb}, at which inequality (9.17) turns into equality (9.3), oscillations cannot originate. The feedback line corresponding to the critical feedback occupies the position OB.

If in an oscillator with inductive feedback and having oscillatory characteristic shown in Fig. 9.6 the coupling M is smoothly increased, then, starting from a critical value M_{cr}, the amplitude of the steady-state oscillation smoothly rises, as shown in Fig. 9.8. Such conditions of self-excitation are called *soft*. From the preceding discussion it is clear that in order to obtain soft excitation conditions, it is necessary that the oscillatory characteristic should go beyond the zero point and have a sufficiently large inclination in the region of low amplitudes. All these requirements are satisfied when using automatic bias.

When forced (external) bias is used, the oscillatory characteristic assumes the shape shown in Fig. 9.9. In this case very strong feedback (line OA, mutual induction M_1) is required for initiating oscillations. After the process of oscillation has been initiated, the coupling can be weakened to a value of M_2 at which the feedback line occupies the position OB. As the coupling is weakened further the oscillations cease. To restore the oscillations, M must be increased to M_1 corresponding to the feedback line OA. Such conditions of self-excitation are called *hard*.

The dependence of the steady-state amplitude $I_{t,st}$ on the value of M under hard excitation conditions is shown in Fig. 9.10, the arrows indicating the direction of variation of M.

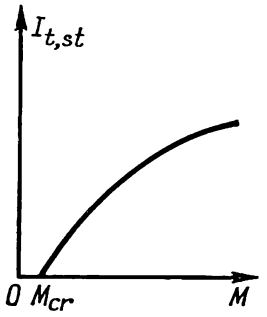


Fig. 9.8. Steady-state amplitude as a fuction of the amount of feedback under soft excitation conditions

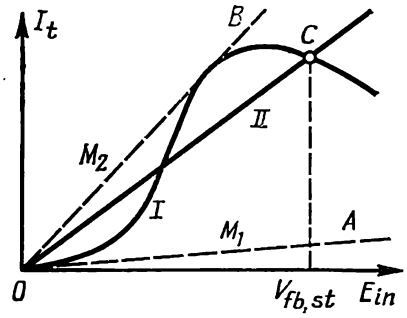


Fig. 9.9. Oscillatory characteristic corresponding to hard excitation conditions

If the external bias is so great that the oscillatory characteristic starts from a point other than zero (Fig. 9.11), no increase in feedback will cause self-oscillation. If oscillations are excited by an

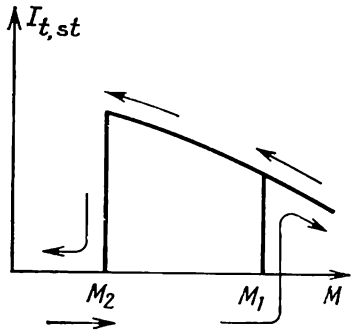


Fig. 9.10. Steady-state amplitude as a function of the amount of feedback under hard excitation conditions

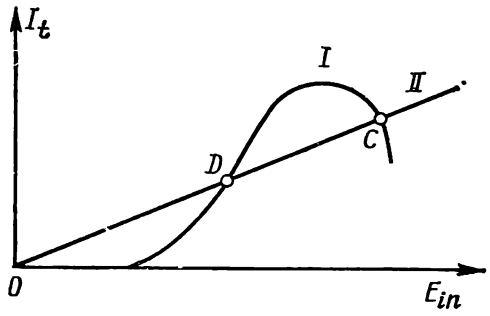


Fig. 9.11. Illustrating the oscillation stability under hard excitation conditions

external force, they may continue to exist after removing this force, provided that feedback is strong enough. Of the two points of intersection of the lines I and II, the point C is stable, while the point D is unstable (here we mean dynamic stability, i.e., oscillation stability). This implies that in the case of small random deviations of the current amplitude in the tank circuit about the point C, the system returns to its original position, while any small deviation of the amplitude in the region of the point D rises progressively and shifts the amplitude I_t either to the stable point C or to the point O (cor-

responding to static stability). The proof of the instability of the point D is analogous to that of the stability of the point C, given in the preceding section.

9.5. EXAMPLES OF SELF-OSCILLATOR CIRCUITS

Figure 9.12 shows three circuit diagrams of single-tuned vacuum-tube self-oscillators differing only in feedback circuits. The circuit with inductive (transformer) feedback (Fig. 9.12*a*) was already dis-

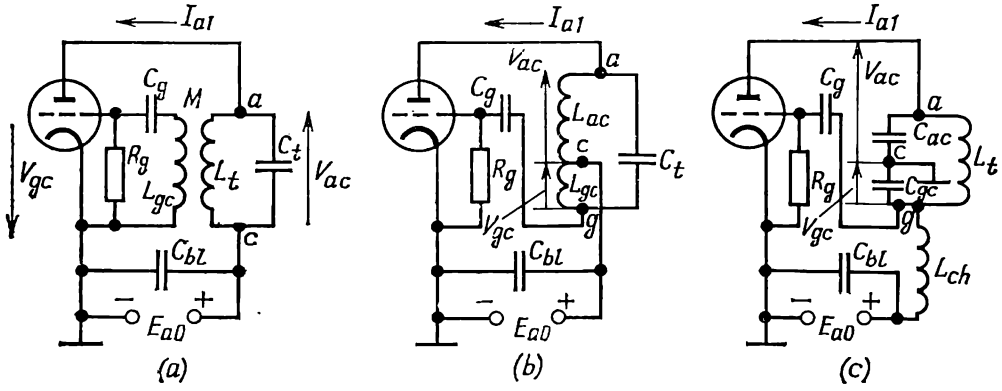


Fig. 9.12. Diagrams of self-oscillators with (*a*) transformer, (*b*) autotransformer, and (*c*) capacitive feedback

cussed in Sec. 9.2 (Fig. 9.3*a*), the only difference being in the method of connection of the resistor R_g (between the grid and cathode). The circuit with conductive (autotransformer) feedback (Fig. 9.12*b*), called the Hartley (tapped-coil) oscillator circuit, and that with capacitive feedback (Fig. 9.12*c*), called the Colpitts (tapped-capacitor) oscillator circuit, often go by the general name impedance-divider oscillator circuits.

In the circuits of Fig. 9.12*a* and *b*, the capacitances C_{bl} of the blocking capacitors are usually so high that for r-f current the points *c* can practically be considered to be connected dead short to the cathode. Since the cathodes are earthed, the points *c* are zero-potential points.

The selective two-port network and the feedback two-port network, shown in Fig. 9.1, in the simple self-oscillators in question are combined in a single tank circuit. The input terminals of this two-port network are the points *a* and *c*, to which are connected the anode and the cathode of the tube, while the output terminals are the points of connection of the grid and cathode (*g* and *c*). Thus, the circuits shown in Fig. 9.12 can be replaced by a single circuit (Fig. 9.13). The power sources are not shown in this diagram.

In the simple circuits under consideration, the oscillation frequency is close to the resonance frequency of the tank circuit. At this frequency the voltage drop $\mathbf{V}_{ac}$ across the tank circuit coincides (or nearly coincides) in phase with the current $\mathbf{I}_{a1}$, while the latter coincides with the voltage $\mathbf{V}_{gc}$. Note that the voltage $\mathbf{V}_{ac}$ is directed, as the current $\mathbf{I}_{a1}$ in the external circuit, from the cathode to the anode. If the voltage at the output of the feedback two-port network is in phase with $\mathbf{V}_{ac}$, it turns to be in antiphase with the initial voltage $\mathbf{V}_{gc}$. From this it follows that the argument of the feedback factor $\mathbf{K}_{fb}(i\omega)$, i.e., the phase shift in the feedback two-port network, must be close to 180°. This conclusion can also be arrived at by reasoning as follows: a single-tuned resonance amplifier shifts the phase of the oscillation being amplified by 180°, therefore, to maintain self-oscillation, the voltage fed from the output to the input through the feedback circuit must get an additional phase shift of 180°. One can easily see how this requirement is met with in the circuits shown in Fig. 9.12. In the circuit with transformer feedback (Fig. 9.12*a*), the 180-degree phase shift is obtained with the proper connection of the coil L_{gc} to the grid and cathode terminals. The magnitude of the feedback factor is

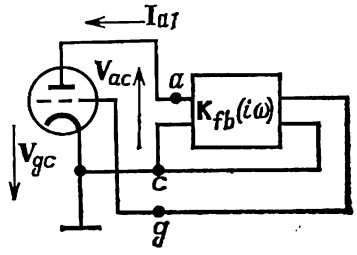


Fig. 9.13. Equivalent circuit of a single-tuned oscillator

$$K_{fb} = M/L_t \tag{9.20}$$

In the autotransformer circuit (Fig. 9.12*b*), the required phasing is achieved by taking the feedback voltage from the inductance coil L_{gc} included into the capacitive branch of the tank circuit. At resonance, the currents in the inductive and capacitive branches of the tank are equal in amplitude and opposite in direction. Consequently, the inductance coils L_{ac} and L_{gc} carry one and the same tank current and form a voltage divider. With respect to the cathode connected to the intermediate point of the tank circuit, the voltages taken from the coils L_{ac} and L_{gc} are in antiphase.

The magnitude of the feedback factor is

$$K_{fb} = L_{gc}/L_{ac} \tag{9.21}$$

The resonance frequency of the tank circuit in the tapped-coil oscillator

$$\omega_r = 1/\sqrt{(L_{ac} + L_{gc})\, C_t} \tag{9.22}$$

Finally, for the circuit with capacitive feedback (Fig. 9.12*c*) we have

$$K_{fb} = C_{ac}/C_{gc} \tag{9.23}$$

and the resonance frequency of the tank circuit

$$\omega_r = 1/\sqrt{L_t C_{eq}} \tag{9.24}$$

where

$$C_{eq} = C_{ac}C_{gc}/(C_{ac} + C_{gc})$$

In the transistor oscillator with capacitive feedback (Fig. 9.14), the external bias voltage V_{BE0} is fed from the voltage divider R_1, R_2 connected to the power-supply source of the collector circuit.

In this system, the feedback factor

$$K_{fb} = C_{CE}/C_{BE} \tag{9.25}$$

and the resonance frequency is defined by a formula similar to (9.24).

When considering all the above circuits, no account has been taken of parasitic factors — interelectrode capacitances, inductances of the tube leads, phase shift of the anode current due to the effect of the electron inertia, etc.

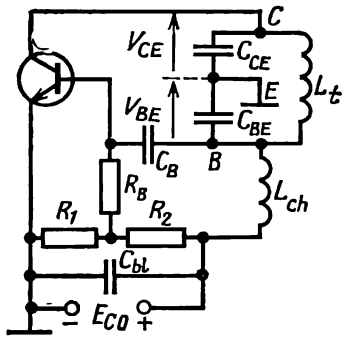


Fig. 9.14. Transistor self-oscillator with capacitive feedback

Therefore, the feedback factor in single-tuned self-oscillators has been found to be independent of frequency. This conclusion is true for relatively low frequencies. With an increase in the operating frequency, the equivalent circuit of the self-oscillator becomes more complex and the feedback factor must be considered with due regard for the above parasitic factors. The frequency dependence of $\mathbf{K}_{fb}$ $(i\omega)$ is especially pronounced in transistor self-oscillators operating at frequencies close to the cutoff frequency of the transistors used. The argument φ_S of the complex slope $\mathbf{S}_{av}$ (see Sec. 9.3) in these oscillators comes to 90° and more. In this case, the argument φ_{fb} of the complex feedback factor $\mathbf{K}_{fb}$ differs from 180°.

Widely used at high frequencies are common-base transistor self-oscillators having some advantages over common-emitter circuits. The circuit diagram of a typical local oscillator used in radio receivers is shown in Fig. 9.15. The automatic bias resistor R_{E0} is inserted into the emitter circuit. Although the above-mentioned frequency dependence of the feedback factor and the complex nature of the

average slope do not affect the configuration of the circuit, the parameters of the tank circuit and of the elements of the feedback

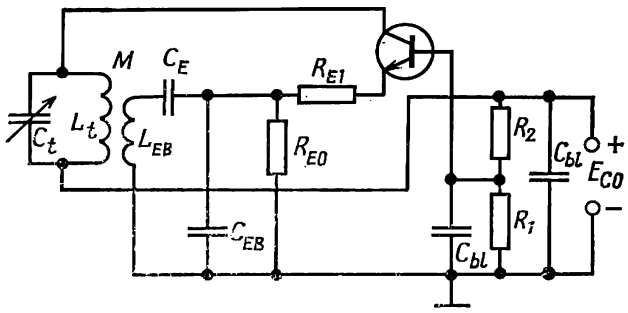


Fig. 9.15. Common-base transistor self-oscillator

circuit differ considerably from those for comparatively low frequencies.

9.6. NONLINEAR EQUATION OF THE SELF-OSCILLATOR

In the preceding sections of this chapter we discussed the conditions under which oscillations originate in the self-oscillator and studied the stability of its steady-state operation. Now it is necessary to consider the whole process of development of self-oscillations, from the moment the self-oscillator is switched on and till steady-state operating conditions have set in. This is very important for a number of applications, where one has to deal with the shaping of short radio pulses (e.g., in pulse radio systems). For complete description of the operation of the self-oscillator, embracing all stages of the transient process, one should abandon the condition of low oscillation amplitude, which forms the basis of linear differential equation (9.9′).

Linear expression (9.6) which was used for setting up the above equation, i.e.,

$$i_a = Sv_{gc} - (v_{ac}/R_i) = S(v_{gc} - Dv_{ac})$$

should be replaced by the nonlinear function

$$i_a = \psi(v_{gc} - Dv_{ac}) \tag{9.26}$$

defining the anode current i_a for any values of v_{gc} and v_{ac}.

For further examination, it is convenient to turn from the diagram shown in Fig. 9.3*a* to that in Fig. 9.16. The parallel equivalent circuit of the tank circuit, used in this diagram, simplifies the setting up of the differential equation for the voltage v_{ac} existing across the tank circuit. With the directions of currents and voltages being as

shown in the figure, we may write the following initial equations:

$$i_C + i_R + i_L = i_a \tag{9.27}$$

$$i_C = C\frac{dv_{ac}}{dt}, \quad i_R = \frac{v_{ac}}{R}, \quad i_L = \frac{1}{L}\int v_{ac}\,dt \tag{9.28}$$

Substituting expressions (9.26) and (9.28) into (9.27), we obtain

$$C\frac{dv_{ac}}{dt} + \frac{1}{R}v_{ac} + \frac{1}{L}\int v_{ac}\,dt = \psi(v_{gc} - Dv_{ac}) \tag{9.29}$$

Now let us transform the argument of the nonlinear function ψ by expressing v_{gc} through v_{ac}. Directly from the circuit with transformer feedback (Fig. 9.16) follows the relation $v_{gc} = M(di_L/dt)$, while from the third equation of (9.28) it follows that $di_L/dt = v_{ac}/L$. Consequently, $v_{gc} = v_{ac}M/L$.

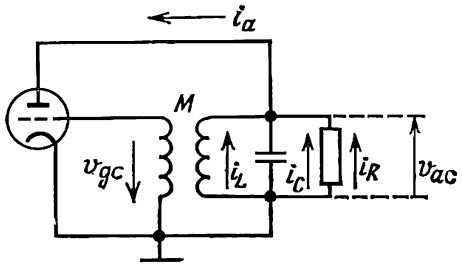


Fig. 9.16. Equivalent circuit of a self-oscillator, which corresponds to equation (9.29)

So,

$$v_{gc} - Dv_{ac} = (M/L - D)\,v_{ac} = (K_{fb} - D)\,v_{ac}$$

Note that if the effect of v_{ac} on the current i_a is not taken account of (for $D \ll K_{fb}$), then $v_{gc} \approx K_{fb}v_{ac}$.

Let us denote

$$K'_{fb} = K_{fb} - D \tag{9.30}$$

Then,

$$\psi(v_{gc} - Dv_{ac}) = \psi[(K_{fb} - D)\,v_{ac}] = \psi(K'_{fb}v_{ac}) \tag{9.31}$$

Substituting this expression into (9.29) and differentiating with respect to t, we obtain

$$\frac{d^2v_{ac}}{dt^2} + \frac{1}{CR}\frac{dv_{ac}}{dt} + \frac{1}{LC}v_{ac} = \frac{1}{C}\frac{d\psi(K'_{fb}v_{ac})}{dt}$$

or

$$\frac{d^2v_{ac}}{dt^2} + \frac{d}{dt}\left[\frac{v_{ac}}{CR} - \frac{1}{C}\psi(K'_{fb}v_{ac})\right] + \frac{1}{LC}v_{ac} = 0 \tag{9.32}$$

As it might be expected, we has obtained a nonlinear equation. Further procedure consists in substitution of any suitable approximation of the function $\psi(K'_{fb}v_{ac})$ into equation (9.32).

Polynomial approximation proves most convenient. So as not to complicate the problem, the incomplete third-degree polynomial [see (8.13)] is usually used:

$$i_a = \psi(K'_{fb}v_{ac}) = a_1K'_{fb}v_{ac} + a_3(K'_{fb}v_{ac})^3 = a_1K'_{fb}v_{ac} - |a_3|(K'_{fb}v_{ac})^3 \tag{9.33}$$

The term $i(V_0)$ entering into expression (8.8) is omitted, for it has no effect on the behaviour of the function v_{ac}. The minus sign before the cubic term is taken in accordance with formula (8.14).

Approximation (9.33) is suitable where the position of the operating point on the current-voltage characteristic curve (at the inflection point, see Fig. 8.5) is fixed. Consequently, no account is taken here of the variation of the bias voltage V_0 in the process of rise of the oscillation amplitude (in the case of automatic bias). None the less, as experience shows, approximation (9.33) makes it possible to reveal the basic features of the process of setting in of oscillations in the self-oscillator operating under soft excitation conditions.

Substituting (9.33) into (9.32), we obtain

$$\frac{d^2v_{ac}}{dt^2}+\frac{d}{dt}\left[\frac{1}{C}\left(\frac{1}{R}-K'_{fb}a_1\right)v_{ac}+|a_3|K'^3_{fb}v^3_{ac}\right]+\frac{1}{LC}v_{ac}=0$$

or

$$\frac{d^2v_{ac}}{dt^2}-(2|\alpha_{eq}|-\gamma_{eq}v^2_{ac})\frac{dv_{ac}}{dt}+\omega_0^2v_{ac}=0 \qquad (9.34)$$

where we have used the notations

$$2\alpha_{eq}=(1/R-K'_{fb}a_1)/C,\quad \gamma_{eq}=3|a_3|K'^3_{fb}/C,\quad \omega_0^2=1/LC \qquad (9.35)$$

It should be noted that in a self-exciting oscillator $\alpha_{eq}<0$ (see Sec. 9.2).

Dividing (9.34) by ω_0^2 and introducing the small parameter

$$\varepsilon=2|\alpha_{eq}|/\omega_0 \qquad (9.36)$$

we obtain

$$\frac{1}{\omega_0^2}\frac{d^2v_{ac}}{dt^2}-\varepsilon\left(1-\frac{\gamma_{eq}v^2_{ac}}{2|\alpha_{eq}|}\right)\frac{1}{\omega_0}\frac{dv_{ac}}{dt}+v_{ac}=0$$

Finally, transforming to dimensionless time $\tau=\omega_0 t$, we obtain the following equation, known as Van der Pol's equation:

$$\frac{d^2v_{ac}}{d\tau^2}-\varepsilon\left(1-\frac{\gamma_{eq}v^2_{ac}}{2|\alpha_{eq}|}\right)\frac{dv_{ac}}{d\tau}+v_{ac}=0 \qquad (9.37)$$

At low voltages, when $\sqrt{\gamma_{eq}/2|\alpha_{eq}|}v_{ac}\ll 1$, equation (9.37) is reduced to a linear equation coinciding with equation (9.9′) [if in the latter we transform to dimensionless time and take into account the difference between the circuits shown in Figs. 9.3 and 9.16]. As the amplitude of the voltage v_{ac} increases, the nonlinearity of the system, due to the quantity $(\gamma_{eq}/2|\alpha_{eq}|)v^2_{ac}$, becomes more noticeable.

No methods are available that would allow the exact solution of nonlinear equation (9.37) to be obtained.

9.7. APPROXIMATE SOLUTION OF THE NONLINEAR EQUATION OF THE SELF-OSCILLATOR

As applied to the present problem, the essence of the approximate method consists in that the solution of the nonlinear equation (9.37) is found in the form of a high-frequency oscillation

$$v_{ac} = A(\omega_0 t)\cos\omega_0 t = A(\tau)\cos\tau \qquad (9.38)$$

where $\omega_0 = \omega_r$ is the resonance frequency of the tank circuit, i.e., $\omega_0 = 1/\sqrt{LC}$, and $A(\tau)$ is a slowly varying function of time.

This approach is based on the assumption that the tank circuit has a high Q-factor, in which case it takes a substantial number of periods for the oscillation amplitude and hence, the energy stored in the circuit to change materially. The condition for the slow variation of the amplitude, specified in Sec. 3.1, is supposed to be satisfied.

Thus, to obtain an approximate solution of equation (9.37), one has to find only the function $A(\tau)$, i.e., the envelope of the oscillation. The oscillation frequency is simply taken at $\omega_0 = 1/\sqrt{LC}$, while the epoch angle which is omitted in solution (9.38), may be anything, depending on the initial conditions of starting the oscillator*.

Let us substitute (9.38) into (9.37), but before that, let us find the derivatives of the function $v_{ac} = A(\tau)\cos\tau$, omitting for the sake of brevity the argument of the function $A(\tau)$ and denoting the derivative $dA(\tau)/d\tau$ as $\dot{A}$:

$$v_{ac} = A\cos\tau, \quad \frac{dv_{ac}}{d\tau} = -A\sin\tau + \dot{A}\cos\tau$$

$$\frac{d^2v_{ac}}{d\tau^2} = -A\cos\tau - \dot{A}\sin\tau - \dot{A}\sin\tau + \ddot{A}\cos\tau \approx -A\cos\tau - 2\dot{A}\sin\tau$$

$$v_{ac}^2\frac{dv_{ac}}{d\tau} = \frac{1}{3}\frac{d(v_{ac}^3)}{d\tau} = \frac{1}{3}\frac{d}{d\tau}(A^3\cos^3\tau)$$

$$= -\frac{A^3}{4}\sin\tau + \frac{1}{4}\frac{dA^3}{d\tau}\cos\tau \approx -\frac{A^2}{4}\sin\tau$$

The term with $\ddot{A}$ is rejected, since the second derivative of a slowly varying function is a quantity of the second order of smallness; the term containing $\cos 3\tau = \cos 3\omega_0 t$, which is obtained when $\cos\tau$ is raised to the third power, is rejected, because the tripled frequency is filtered out by the tank circuit tuned to the frequency ω_0. In addition, it should be borne in mind that the last term $\frac{1}{4}(dA^3/d\tau)\cos\tau$,

* In fact, the phase and consequently, the oscillation frequency in the process of setting in of oscillations may depend on time. A second or even higher-order approximations are necessary for finding the frequency correction.

after being substituted into equation (9.37) and multiplied by the small quantity ε, is also rejected as being negligible in comparison with the terms having the coefficients A, $\dot{A}$ or εA^3. The product $\varepsilon\dot{A}\cos\tau$ is also rejected. As a result, equation (9.37) is reduced to

$$2\dot{A}\sin\tau - \varepsilon A\sin\tau + \frac{\varepsilon\gamma_{eq}}{2\,|\,\alpha_{eq}\,|}\,\frac{A^3}{4}\sin\tau$$

$$= \left(2\dot{A} - \varepsilon A + \frac{\varepsilon\gamma_{eq}}{2\,|\,\alpha_{eq}\,|}\,\frac{A^3}{4}\right)\sin\tau = 0$$

Since $\sin\tau \neq 0$, we come to the following equation for the amplitude:

$$2\dot{A} - \varepsilon\left(A - \frac{\gamma_{eq}}{2\,|\,\alpha_{eq}\,|}\,\frac{A^3}{4}\right) = 0$$

Multiplying this equation by A and taking into account that $2A\dot{A} = dA^2/d\tau$, we rewrite this equation in the form

$$\frac{dA^2}{d\tau} - \varepsilon\left(A^2 - \frac{\gamma_{eq}}{2\,|\,\alpha_{eq}\,|}\,\frac{A^4}{4}\right) = 0 \tag{9.39}$$

Thus, we have obtained a first-degree equation in the unknown square of the amplitude. The steady-state amplitude A_{st} is found by equating to zero the derivative of A^2. Thus, we have

$$A_{st}^2 - \frac{\gamma_{eq}}{2\,|\,\alpha_{eq}\,|}\,\frac{A_{st}^4}{4} = 0$$

whence

$$A_{st} = 2/\sqrt{\gamma_{eq}/2\,|\,\alpha_{eq}\,|} \tag{9.40}$$

and equation (9.39) assumes the form

$$dA^2/d\tau - \varepsilon\,(A^2 - A^4/A_{st}^2) = 0 \tag{9.41}$$

This equation is solved by using the substitution $A^2 = 1/x$. Then, $A_{st}^2 = 1/x_{st}$ and

$$\frac{d\,(1/x)}{d\tau} - \varepsilon\left(\frac{1}{x} - \frac{x_{st}}{x^2}\right) = -\frac{1}{x^2}\,\frac{dx}{d\tau} - \varepsilon\left(\frac{1}{x} - \frac{x_{st}}{x^2}\right) = 0$$

or

$$\frac{dx}{d\tau} + \varepsilon\,(x - x_{st}) = 0$$

Separating the variables

$$dx/(x - x_{st}) = -\varepsilon d\tau$$

and integrating, we get

$$\ln\,(x - x_{st}) + C = -\varepsilon\tau$$

Let the initial value of the oscillation amplitude at time $\tau = 0$ be equal to A_0. Then the value of x_0 corresponding to this instant is

equal to $1/A_0^2$. Setting in the last expression $\tau = 0$, we find the integration constant C:

$$C = -\ln (x_0 - x_{st})$$

Hence,

$$\ln (x - x_{st}) - \ln (x_0 - x_{st}) = \varepsilon\tau$$

and

$$\frac{x - x_{st}}{x_0 - x_{st}} = e^{\varepsilon\tau} \tag{9.42}$$

whence

$$x = x_{st} [1 + (x_0/x_{st} - 1)\, e^{\varepsilon\tau}]$$

or

$$\frac{1}{A^2} = \frac{1}{A_{st}^2}\left[1 + \left(\frac{A_{st}^2}{A_0^2} - 1\right) e^{\varepsilon\tau}\right]$$

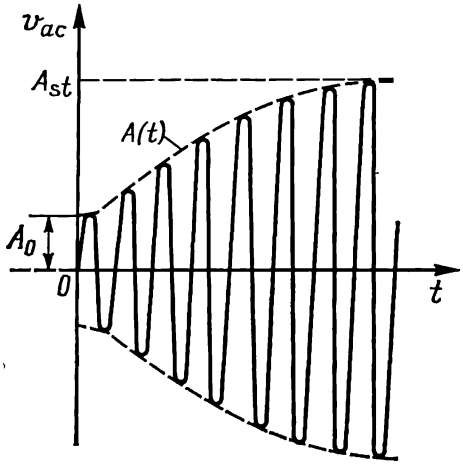


Fig. 9.17. Process of setting in of oscillations in a self-oscillator

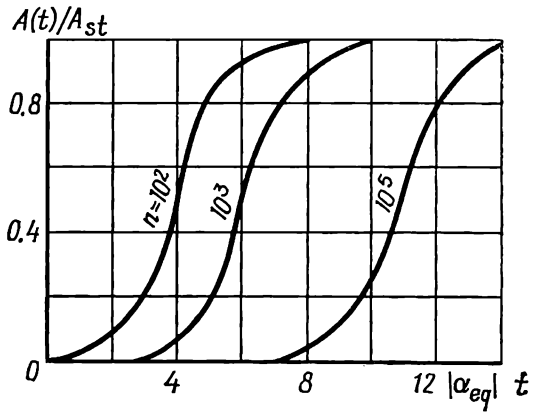


Fig. 9.18. Character of rise of the self-oscillation envelope under different initial conditions

Finally, taking into account expression (9.36) and $\tau = \omega_0 t$ and also the fact that $\alpha_{eq} < 0$, we get

$$A = \frac{A_{st}}{\sqrt{1 + (A_{st}^2/A_0^2 - 1)\, e^{\varepsilon\tau}}} = \frac{A_{st}}{\sqrt{1 + (A_{st}^2/A_0^2 - 1)\, e^{-2|\alpha_{eq}|\, t}}} \tag{9.43}$$

Substituting this expression into (9.38), we have

$$v_{ac}(t) = \frac{A_{st}}{\sqrt{1 + (A_{st}^2/A_0^2 - 1)\, e^{-2|\alpha_{eq}|\, t}}} \cos (\omega_0 t + \theta_0) \tag{9.44}$$

where θ_0 is the epoch angle depending on the oscillator starting conditions.

As a rule, $A_{st}/A_0 \gg 1$. Therefore, at low values of $2|\alpha_{eq}|t$, the denominator

$$\sqrt{1+\left(\frac{A_{st}^2}{A_0^2}-1\right)e^{-2|\alpha_{eq}|t}} \approx \frac{A_{st}}{A_0}e^{-|\alpha_{eq}|t}$$

and expression (9.44) assumes the form

$$v_{ac}(t) \approx A_0 e^{|\alpha_{eq}|t}\cos(\omega_0 t+\theta_0)$$

coinciding with the form of expression (9.10) derived for *linear* operating conditions (for small amplitudes).

When $t \to \infty$ (steady-state conditions)

$$v_{ac}(t) = A_{st}\cos(\omega_0 t+\theta_0)$$

The amplitude limitation due to the insertion of the cubic term into approximation (9.33) of the current-voltage characteristic is illustrated in Fig. 9.17 representing the process of setting in of oscillations in the self-oscillator.

The character of variation of the envelope $A(t)/A_{st}$ for several values of the parameter $n = A_{st}/A_0$ is shown in Fig. 9.18.

From expression (9.39) and Fig. 9.18 it is clear that the setting-in time of steady-state conditions depends essentially on the initial amplitude, i.e., on the initial conditions of starting the oscillator. This is very important for oscillators operating under pulse conditions.

In conclusion, it should be noted that for satisfactory description of the process of setting in of oscillations under hard excitation conditions, at least the fifth-power term is required in polynomial (8.8).

9.8. SELF-OSCILLATORS WITH INTERNAL FEEDBACK

When considering the mechanism of origination of oscillations in an oscillator (see Sec. 9.2), we have met with the notion of negative resistance that is introduced into the tank circuit if the phase of feedback is selected properly. In this case, in accordance with the generalized diagram of a self-oscillating system (see Fig. 9.1), we have meant external feedback.

There are, however, some electron devices which allow one to obtain negative resistance by using the falling sections of their current-voltage characteristics without introducing special feedback elements into the system. These devices may be exemplified by a tunnel diode or ordinary tetrodes and pentodes operating with properly chosen electrode voltages.

Figure 9.19*a* shows the current-voltage characteristic of a tunnel diode, representing the dependence of the forward current of the diode on the positive bias voltage. In the falling section *a-b* the differential resistance of the diode is negative:

$$R_- = dv/di = \cot \gamma$$

where γ is the slope angle of the tangent to the curve $i = f(v)$ at the operating point V_0.

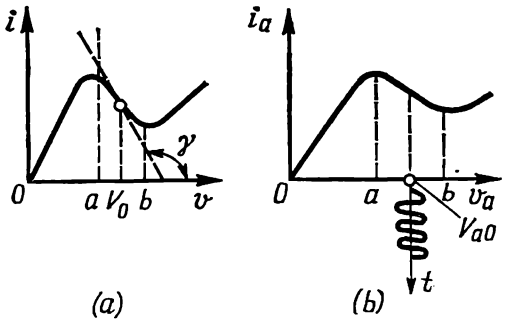


Fig. 9.19. Current-voltage characteristics which make it possible to realize negative resistance
(*a*) a tunnel diode; (*b*) a tetrode with dynatron effect

The characteristic of a tetrode with pronounced dynatron effect has a similar shape (Fig. 9.19*b*).

If an electron device having such a current-voltage characteristic is connected to an oscillatory circuit, one can effect generation of r-f oscillations. In this case, one will obtain a self-oscillator with *internal feedback*. Figure 9.20*a* shows a diagram of a dynatron oscillator in which a tetrode is used as a negative resistance. This system has no special feedback elements. The negative character of the tube resistance is achieved by setting the operating point on the falling section of the characteristic curve (Fig. 9.19*b*). For this purpose, the anode is fed with a voltage E_{a0} which is lower than that applied to the screen grid. The equivalent circuit of a circuit shunted by a negative resistance R_- is shown in Fig. 9.20*b*. With respect to this resistance, v_a is regarded as an electromotive force, so that the current i_a and the voltage v_a are related by the relation $i_a = v_a/R_-$.

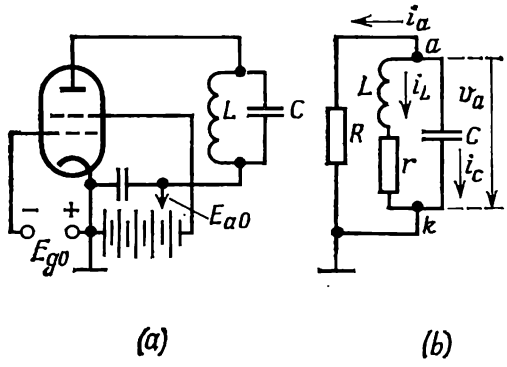


Fig. 9.20. (*a*) Dynatron oscillator and (*b*) its equivalent circuit

Examining, as in Sec. 9.2, the initial stage of starting oscillation (low amplitudes), we shall use equations similar to equations (9.4) and (9.5), the only difference being that the first equation of (9.4) will be written in the form $i_L + i_C = -i_a$.

Then, taking into account the above expression $i_a = v_a/R_-$ and using the third equation of (9.4), we obtain, instead of (9.8), the

following linear differential equation

$$\frac{d^2 i_L}{dt^2} + \left(\frac{r}{L} + \frac{1}{CR_-}\right)\frac{di_L}{dt} + \frac{1 + r/R_-}{LC} i_L = 0 \qquad (9.45)$$

To make the oscillation amplitude rise, the coefficient multiplying the first derivative must be negative. From this we obtain the

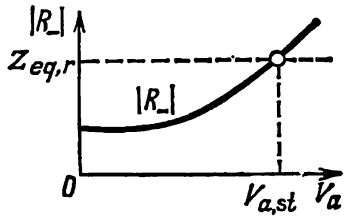


Fig. 9.21. Determining the steady-state oscillation amplitude in a self-oscillator with internal feedback

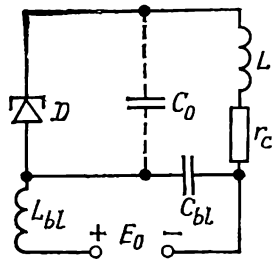


Fig. 9.22. Oscillator built around a tunnel diode

condition for origination of oscillations in the form

$$\frac{r}{L} - \frac{1}{C}\left|\frac{1}{R_-}\right| < 0$$

or

$$|R_-| < \frac{L}{rC} = Z_{eq,\,r} = \rho Q = Q/\omega_r C \qquad (9.46)$$

where $|R_-|$ is the magnitude of the negative resistance, $Z_{eq,\,r}$ is the resonant impedance of the tank circuit, Q is the quality factor, and $\rho = \sqrt{L/C}$ is the characteristic impedance of the tank circuit.

When the rasistance $|R_-|$, that depends on the oscillation amplitude (when operating on the nonlinear portion of the tube characteristic), increases to

$$|R_-(V_a)| = Z_{eq,\,r} \qquad (9.47)$$

the oscillation amplitude in the self-oscillator reaches its steady state. The conditions are stable if at the point of intersection of the horizontal $Z_{eq,\,r}$ the curve $|R_-(V_a)|$ has a positive slope (Fig. 9.21). What was said in the preceding sections about the character of the nonlinear dependence of the average slope on the amplitude of the control voltage can in this case be extended to cover the character of the dependence of the inverse of $|R_-|$ on the voltage V_a.

Because of the insufficient stability of the dynatron effect and low efficiency, the dynatron oscillators are used rather seldom. Oscillators built around a tunnel diode (Fig. 9.22) have gained a much wider application. The blocking choke L_{bl} and the blocking

capacitor C_{bl} ($C_{bl} \gg C_0$) protect the d-c circuit from radio-frequency current; C_0 is the intrinsic capacitance of the diode; r_c is the resistance accounting for the losses in the diode crystal and tank elements. The advantage of the tunnel diode consists in its low negative resistance (about 10 to 100 Ω). In spite of the comparatively high intrinsic capacitance of the diode (several tens of picofarads), self-excitation condition (9.46) is satisfied within a fairly wide frequency range (up to VHF).

For example, with the quality factor of the tank circuit, Q, equal to 50, the intrinsic capacitance of the diode, C_0, equal to 50 pF, and $|R_-| = 50\ \Omega$, the maximum oscillation frequency is $f_{max} \approx \approx Q/2\pi\,|R_-|\,C_0 \approx 3 \times 10^9$ Hz = 300 MHz.

9.9. SELF-OSCILLATOR WITH A DELAY LINE IN THE FEEDBACK CIRCUIT

Assume that we have a self-oscillator with a selective load and a delay line in the feedback loop. Such an oscillator can be represented in the form of a generalized diagram (Fig. 9.23) similar

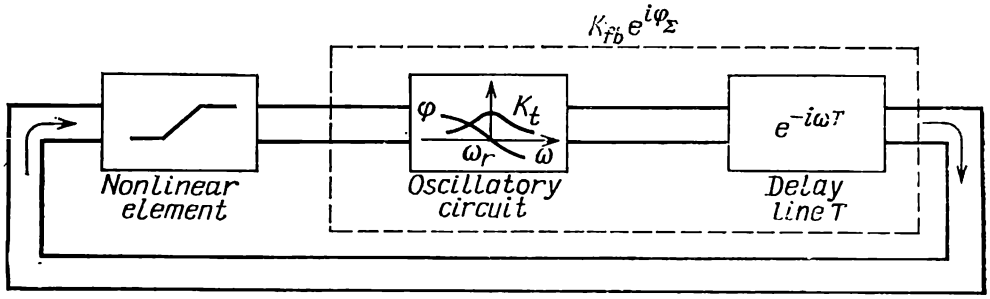


Fig. 9.23. Self-oscillator with a delay line in the feedback loop

to that shown in Fig. 9.1. Regarding the delay line as an ideal two-port network with a transfer function $e^{-i\omega T}$, we may represent the linear part of the oscillator circuit, consisting of a tank circuit and a delay T, in the form of a single feedback two-port network with a transfer function

$$\mathrm{K}_{fb}\,[i\,(\omega - \omega_r)] = K_t\,(\omega - \omega_r)\,e^{i\varphi_t}e^{-i\omega T} = K_{fb}e^{-i\varphi_\Sigma} \qquad (9.48)$$

where K_t is the magnitude of the transfer function of the tank circuit whose resonance frequency is ω_r, and φ_t is the phase characteristic of the tank circuit. Within the passband of the circuit, we may assume, that $\varphi_t \approx -(\omega - \omega_r)\,\tau_t$, where τ_t is the time constant of the tank circuit.

The introduction of the delay line in the circuit does not change the magnitude of the transfer function, but has a significant effect

on the resultant phase characteristic

$$\varphi_\Sigma = -(\omega - \omega_r)\,\tau_t - \omega T \qquad (9.49)$$

With an adequately large delay T, the slope of the resultant phase characteristic is defined mainly by the term ωT, it being quite possible that within the passband of the oscillatory circuit the change in φ_Σ is so large that it exceeds several complete revolutions of 2π. Such a case is shown in Fig. 9.24 where ω_{-2}, ω_{-1}, ω_1, ω_2, . . . are the frequencies lying within the passband of the tank circuit,

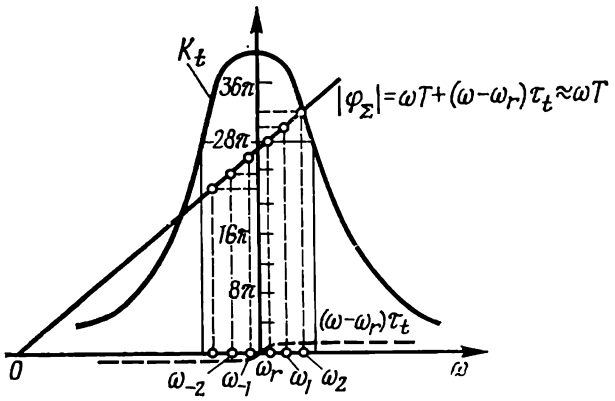


Fig. 9.24. Location of possible oscillation frequencies for a self-oscillator with a delay line inserted into the feedback loop

for which the ordinates of the phase characteristic φ_Σ are equal to $n2\pi$, where n is an integer. Since for the above frequencies the balance of phases and amplitudes is ensured (see Sec. 9.3), any one of them can be the frequency of self-oscillation. The introduction of a sufficiently large delay into the feedback loop imparts the system a *multifrequency character*. In this case, the role of the tank circuit is reduced to limiting the number of frequencies for which the amplification required for self-oscillation is ensured.

A question arises whether several self-oscillations of different frequencies can exist simultaneously. The answer to this question depends on such factors as the number of frequencies in the passband of the circuit, for which the phase balance is ensured, the shape of the amplitude-frequency characteristic of the selective load, self-excitation conditions (soft or hard), and some others.

First, let us consider the case where the passband of a solitary tank circuit contains only two frequencies ω_1 and ω_2 for which self-oscillation is possible. This means that in the band $2\Delta\omega_0 = 2/\tau_t$ (τ_t is the time constant of the tank circuit) the phase shift in the delay line T approaches 2π, i.e., $2\Delta\omega_0 T \approx 2\pi$ or $T \approx \pi/\Delta\omega_0 \approx \pi\tau_t$.

The location of ω_1 and ω_2 on the frequency axis is shown in Fig. 9.25. E_1 and E_2 stand for the amplitudes of oscillations of the given frequencies at some instant of time after starting the oscillator operating under soft excitation conditions. While current is circulating through the closed feedback loop, the relation between the amplitudes E_1 and E_2 changes in favour of E_1 during each passage through the nonlinear element. The same situation was met with when we discussed the suppression of a weak signal in the amplitude limiter in Sec. 8.9. As a result, the oscillation of frequency ω_2 is completely suppressed and only one oscillation—that of frequency ω_1 remains in the system, since the starting conditions were more favourable for this oscillation.

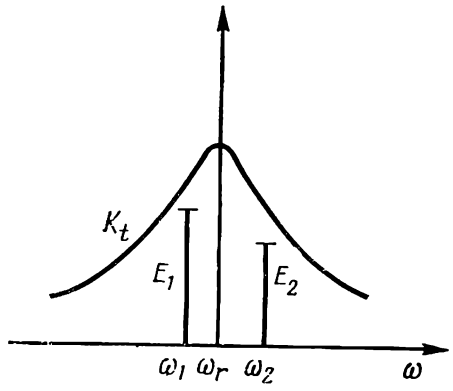


Fig. 9.25. Suppression of a weak oscillation in a self-oscillator with a delay line in the feedback loop

It is different with the self-oscillator operating under hard self-excitation conditions, where an external source of oscillations is required to start self-oscillations. Depending on the choice of the starting frequency, stationary conditions can be established in the oscillator at any of the two frequencies: ω_1 or ω_2. From this it is clear that "hard" self-oscillator with a delayed feedback can be used as a device for storing one of the several frequencies fed to it at the starting moment.

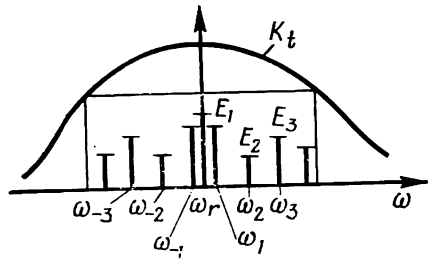


Fig. 9.26. Shape of the amplitude-frequency characteristic of a selective circuit, which is unfavourable for simultaneous generation of several frequencies

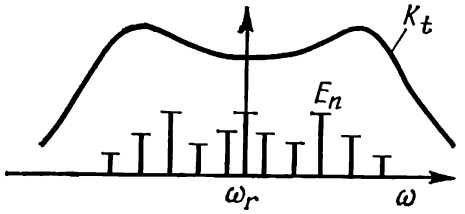


Fig. 9.27. Shape of the amplitude-frequency characteristic of a selective circuit, which permits of stable generation of several frequencies simultaneously

Now let us return to the self-oscillator operating under soft self-excitation conditions and assume that within the passband of the tank circuit there are a number of possible oscillation frequencies.

Since these frequencies are equally spaced along the ω-axis (Fig. 9.26), we may assume that there exists a combination of oscillations with frequencies ω_1, ω_2, ω_3, . . ., whose amplitude and phase relations are typical of angle modulation. Such a complex oscillation of constant amplitude passes through a nonlinear device (amplitude limiter) without deformation, i.e., without any change in the relations between the individual components of the spectrum. This means that the nonlinear part of the self-oscillator does not impede the simultaneous generation of a set of frequencies. However, this is not sufficient for stable generation. It is necessary that the transfer function of the selective circuit should provide for the preservation of the interspectral relations. The amplitude-frequency characteristic shown in Fig. 9.26 does not satisfy this condition. A more detailed analysis [7] shows that for stable generation of a spectrum of frequencies, the amplitude-frequency characteristic of the oscillatory system must have a saddle-shaped nonuniformity (Fig. 9.27).

The self-oscillator with a delayed feedback has some other interesting features that are due to the steep phase characteristic, e.g., high stability of the oscillation frequency.

9.10. ACTION OF A HARMONIC E.M.F. ON CIRCUITS WITH POSITIVE FEEDBACK. REGENERATION

In radio engineering, the term regeneration implies compensation for losses in an oscillatory circuit with the aid of positive feedback. Regeneration can also be used for amplification of oscillations. The circuit diagram of a regenerative amplifier, shown in Fig. 9.28, externally does not differ from that of the self-oscillator. However, in the regenerative amplifier the amount of feedback is not brought to the value corresponding to the self-oscillation threshold. In this case, the circuit losses are compensated for only partially and the action of feedback is reduced solely to increasing the Q-factor of the tuned circuit.

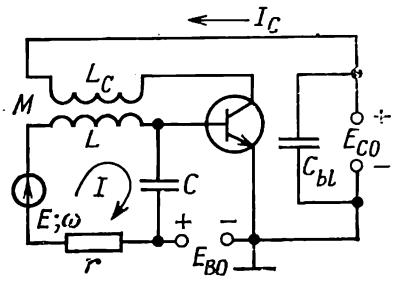


Fig. 9.28. Regenerative amplifier

In the circuit shown in Fig. 9.28, the amplified oscillation with an amplitude E and frequency ω is introduced into the L-C-r tank while feedback is effected by means of the coil L_C through which flows the a-c component I_C of the collector current and which is inductively coupled with the tank coil L. The effect of the parameters of the transistor and feedback on the equivalent parameters of the tuned circuit is easy to reveal. First, let us consider a simple case of low amplitudes, where use is made of a small portion of the

transistor characteristic and the average slope S_{av} can be taken equal to the slope S.

With a harmonic e.m.f. of complex amplitude $\mathbf{E}$ and frequency ω at the amplifier input, the amplitude of the current in the tank circuit (under steady-state conditions) can be found using the following expression:

$$\mathbf{I} = (\mathbf{E} + \mathbf{V}_{fb})/[r + i(\omega L - 1/\omega C)] \tag{9.50}$$

where $\mathbf{V}_{fb}$ is the complex amplitude of the feedback voltage applied to the tank circuit. Since the amplifier circuit is assumed to be linear (for low amplitudes), this voltage coincides in frequency with the e.m.f. $\mathbf{E}$ and may differ from the latter only in phase.

It is obvious that

$$\mathbf{V}_{fb} = i\omega M \mathbf{I}_C$$

and the amplitude $\mathbf{I}_C$ of the a-c component of the collector current (if the output circuit reaction is disregarded) is given by

$$\mathbf{I}_C = S\mathbf{V}_C = S\mathbf{I}/i\omega C$$

where $\mathbf{V}_C$ is the amplitude of the voltage across the tank capacitor.

Thus, we may write

$$\mathbf{V}_{fb} = i\omega MS\mathbf{I}/i\omega C = MS\mathbf{I}/C$$

Substituting this expression into equation (9.50) and solving it for $\mathbf{I}$, we obtain

$$\mathbf{I} = \frac{\mathbf{E}}{r - MS/C + i(\omega L - 1/\omega C)} \tag{9.51}$$

As it might be expected (see Sec. 9.2), the effect of positive feedback is reduced to decreasing the loss resistance in the tank circuit by an amount equal to the negative resistance

$$r_{neg} = MS/C \tag{9.52}$$

Thus, the equivalent resistance of the tank circuit with positive feedback is

$$r_{eq} = r - r_{neg} = r - MS/C \tag{9.53}$$

and the Q-factor is given by

$$Q_{eq} = \frac{\rho}{r_{eq}} = \frac{\rho}{(r - MS/C)} = \frac{\rho/r}{1 - \frac{MS/C}{r}} \tag{9.54}$$

The ratio

$$Q_{eq}/Q_0 = 1\Big/\left(1 - \frac{MS/C}{r}\right) = 1\Big/\left(1 - \frac{r_{neg}}{r}\right) \tag{9.55}$$

where $Q_0 = \rho/r$ is the Q-factor of the tank circuit without feedback and can be regarded as the amplification of the system at the resonance frequency $\omega = \omega_r$, when $\omega L - 1/\omega C = 0$.

By increasing M, it is possible to substantially increase Q_{eq} and hence, obtain a higher amplification. It should be borne in mind,

however, that increasing Q_{eq} at a given constant resonance frequency ω_r of the tank circuit narrows its transmission band. Therefore, the maximum permissible value of MS/Cr can be found from the inequality

$$\frac{MS}{Cr} < 1 - \frac{1}{Q_{eq}/Q_0} = \frac{Q_{eq} - Q_0}{Q_{eq}} \tag{9.56}$$

As $MS/Cr \to 1$, the circuit becomes unstable, and with $MS/Cr > 1$ self-oscillations arise, i.e., the amplifier turns into an oscillator.

It is different with high amplitudes of the voltage at the amplifier input. As the amplitude across the base-to-emitter circuit is increased, the slope S_{av} decreases and, according to (9.52), so does r_{neg}. The dependence of S_{av} and r_{neg} on the amplitude causes nonlinear distortion of the signal being amplified.

Substituting S_{av} for S in expression (9.51), we obtain

$$\mathrm{I} = \frac{\mathrm{E}}{r - MS_{av}/C + i(\omega L - 1/\omega C)} = \frac{\mathrm{E}}{r_{eq} + i(\omega L - 1/\omega C)} \tag{9.57}$$

and at resonance

$$I_{max} = E/r_{eq} \tag{9.58}$$

where

$$r_{eq} = r - MS_{av}/C \tag{9.59}$$

From expression (9.58) it follows that when amplifying an amplitude-modulated signal, i.e., when $E = E(t)$, the current amplitude $I(t)$ will vary according to a law differing from the modulation law of $E(t)$ (because of the dependence of S_{av} on the oscillation amplitude). Thus, nonlinear distortion of the amplified signal occurs, which is accompanied by the formation of new frequencies.

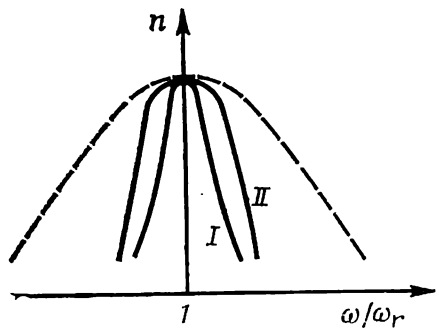


Fig. 9.29. Resonance characteristics of a regenerative loop

I — for low amplitudes; *II*—for high amplitudes; dashed line—without regeneration

The nonlinear character of the circuit in consideration (for high amplitudes of E) also affects the shape of the resonance characteristic of the tuned circuit with positive feedback. In fact, when the frequency ω of the input signal deviates from the resonance frequency ω_r, the reactance $x = \omega L - 1/\omega C$ in (9.57) rises, this resulting in a decrease in the amplitude of the current I. But the decrease in I in turn leads to a decrease in r_{eq} [see expression (9.59)], which to some extent compensates for the effect of the rise of x. As a result, the upper portion of the resonance characteristic becomes flattened and the more so as the amplitude of the external e.m.f. acting on

the circuit grows higher. When the detuning is substantial the reactance has a dominating effect on the amplitude, and the resonance characteristic quickly drops almost to zero (Fig. 9.29).

From this it is clear that regeneration can be used effectively for the amplification of weak signals only.

The phenomenon of regeneration has often to be dealt with in radio engineering. Sometimes regeneration occurs in amplifying devices as a result of a parasitic feedback and this may lead to signals distortion.

9.11. ACTION OF A HARMONIC E.M.F. ON A SELF-OSCILLATOR. FREQUENCY PULLING

The behaviour of a self-oscillator acted upon by an external force depends essentially on the amplitude and frequency of this force.

If the excitation amplitude is low compared to the self-oscillation amplitude and, at the same time, the excitation frequency ω differs substantially from the frequency $\omega_{os} = \omega_r$ of the free-running self-oscillator (ω_r is the resonance frequency of the tank circuit of the oscillator) the action of the external e.m.f. is reduced to a modulation effect which manifests itself in the variation of the self-oscillation amplitude and phase according to a very complex law. As the frequency ω approaches ω_r, the picture changes. The oscillation frequency ω_{os} is "pulled" to the frequency ω of the external e.m.f. and at some value of $\Delta\omega = \omega - \omega_r$ depending on the amplitude ratio the self-oscillator starts operating exactly at the frequency $\omega_{os} = \omega$ without any signs of modulation. The oscillation frequency is "pulled" by the frequency of the driving e.m.f.

The phenomenon of frequency pulling is widely used in a number of radio engineering devices where it is necessary to effect forced synchronization of a self-oscillator by means of a low-power source of oscillations. In some cases, if there is a parasitic coupling between two self-oscillators, the phenomenon of frequency pulling occurs spontaneously and hinders the independent operation of these oscillators at close frequencies.

Let us consider the mechanism of frequency pulling in a simple single-tuned self-oscillator with transformer feedback when an independent source of e.m.f. is connected in series into the base-to-emitter circuit (Fig. 9.30). It should be emphasized that such a diagram has been selected for the sake of definiteness of argumentation only. As to the determination of general relations, the type of self-oscillators circuit and the method of application of the driving e.m.f. are of no principal importance. The oscillation frequency (in the absence of external e.m.f.) is taken to be equal to the resonance frequency $\omega_r = 1/\sqrt{LC}$ of the tuned circuit of the oscillator.

Let us first consider the phase balance in the self-oscillator acted upon by an external e.m.f. $e(t) = E\cos(\omega t + \theta_0)$, assuming that frequency pulling is steady, i.e., the oscillation frequency ω_{os} is equal to the frequency ω differing from the resonance frequency ω_r of the tank circuit, the amplitude E being considered to be so small that all the basic self-oscillation parameters—the amplitude of the first harmonic of the collector current, I_1, the amplitude of the

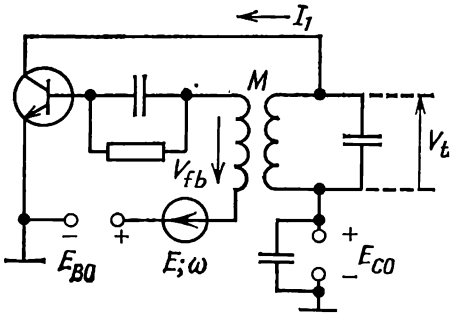


Fig. 9.30. Self-oscillator with a synchronizing e.m.f. source in the base-to emitter circuit

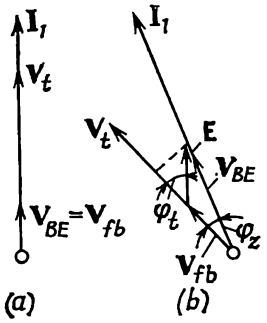


Fig. 9.31. Vector diagrams of voltage and currents in a self-oscillator
(*a*) without external action; (*b*) under conditions of frequency pulling

voltage V_t across the tank circuit, and the amplitude of the feedback voltage V_{fb}—remain the same as in the absence of the drive. In other words, the drive changes only the phase relations in the self-oscillator.

Before switching on the source of the external e.m.f., these relations are characterized by the vector diagram shown in Fig. 9.31*a*. The current $\mathbf{I}_1$ is in phase with the voltage $\mathbf{V}_{BE} = \mathbf{V}_{fb}$, while the voltage $\mathbf{V}_t$ is in phase with the current $\mathbf{I}_1$ (the voltage across the tuned circuit is read from the emitter to the collector).

The epoch angle of the current $\mathbf{I}_1$ is selected arbitrarily, since in the self-oscillator the phase of the self-oscillation depends on the initial starting conditions. After the external e.m.f. $e(t) = E\cos\omega t$ is switched on (the epoch angle θ_0 is taken at zero) and steady-state conditions have set in, the diagram assumes the form shown in Fig. 9.31*b*. This diagram is plotted on the basis of the following conditions:

(a) between the current $\mathbf{I}_1$ and the voltage $\mathbf{V}_t$ there is a phase shift φ_z depending on the detuning of the tank circuit with respect to the oscillation frequency ω. Setting for the sake of definiteness $\omega < \omega_r$, we come to a conclusion that the vector $\mathbf{V}_t$ must lead the vector $\mathbf{I}_1$ by the angle

$$\varphi_z = \arctan\left[\frac{2(\omega-\omega_r)Q}{\omega_r}\right] \qquad (9.60)$$

where Q is the quality factor of the tank circuit;

(*b*) the current $\mathbf{I}_1$ is in phase with the resultant voltage $\mathbf{V}_{BE}$;

(c) the feedback voltage $\mathbf{V}_{fb}$, that is related to the tank voltage $\mathbf{V}_t$ by the relation $\mathbf{V}_{fb} = M\mathbf{V}_t/L$, is independent of frequency. Therefore, the directions of the vectors $\mathbf{V}_{fb}$ and $\mathbf{V}_t$ coincide.

From the diagram it is clear that the disturbance of the phase balance of the self-oscillator in the collector circuit by the angle φ_z (in the lead direction) due to the detuning of the tank circuit (for $\omega < \omega_r$) is compensated for by the fact that the resultant voltage $\mathbf{V}_{BE}$ across the base-to-emitter circuit lags behind $\mathbf{V}_{fb}$ by the angle φ_z.

When $\omega > \omega_r$, the phase shift in the collector circuit is lagging, while in the base-to-emitter circuit it is leading.

From conditions (*b*) and (*c*) and directly from the diagram* in Fig. 9.31*b* we have the following expression:

$$\sin \varphi_z = E \sin \varphi_t / V_{BE} \tag{9.61}$$

where φ_t is the phase shift between $\mathbf{E}$ and $\mathbf{V}_t$.

Thus, if frequency pulling conditions actually exist, equalities (9.60) and (9.61) are satisfied simultaneously.

Proceeding from the above-stipulated condition that E is small in comparison with V_{fb}, we may consider that $V_{BE} \approx V_{fb}$ and that

$$\sin \varphi_z \approx \varphi_z \approx (E/V_{BE}) \sin \varphi_t \approx (E/V_{fb}) \sin \varphi_t \tag{9.62}$$

Since the angle φ_z is small, we may replace the tangent in expression (9.60) by its argument

$$\tan \varphi_z \approx \varphi_z = 2 (\omega - \omega_r) Q/\omega_r \tag{9.63}$$

Equating the right-hand sides of the last two expressions, we get the following important relation:

$$2 (\omega - \omega_r)/\omega_r = (E/V_{fb}) \sin \varphi_t (1/Q) \tag{9.64}$$

From this relation it follows that with the given difference between the frequencies ω and ω_r, the phase shift of the voltage $\mathbf{V}_t$ relative to the synchronizing oscillation is

$$\varphi_t = \arcsin \left[\frac{2 (\omega - \omega_r) Q/\omega_r}{E/V_{fb}} \right] \tag{9.65}$$

Relations (9.64) and (9.65) make sense, provided the absolute detuning $| \omega - \omega_r |$ does not exceed some maximum value at which $| \sin \varphi_t | = 1$. From physical considerations it is clear that this maximum value $| \omega - \omega_r |_{\max}$ corresponds to the boundaries of the pulling range. Substituting $\sin \varphi_t = \pm 1$ into (9.64), we find

* This diagram is preferably plotted proceeding from an arbitrarily selected position of the vector $\mathbf{I}_1$; after that, the whole diagram must be turned through an angle at which the position of the vector $\mathbf{E}$ corresponds to the given epoch angle θ_0 of the external e.m.f.

the total relative width of the pulling range in the form

$$\frac{2\Delta\omega_{max}}{\omega_r} = \frac{2\,|\,\omega-\omega_r\,|_{max}}{\omega_r} \approx \left(\frac{E}{V_{fb}}\right)\left(\frac{1}{Q}\right) = \left(\frac{E}{V_{fb}}\right) d \qquad (9.66)$$

Thus, the pulling range is proportional to the ratio between the amplitude of the external e.m.f. and the oscillation amplitude of the free-running self-oscillator and to the attenuation $d = 1/Q$ of the tank circuit.

Where the external e.m.f. is applied directly to the tank circuit of the self-oscillator, the expression for the pulling range may be

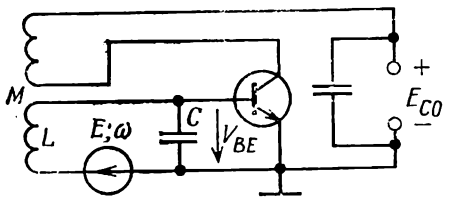


Fig. 9.32. Connection of a synchronizing e.m.f. source into the tank circuit of a self-oscillator

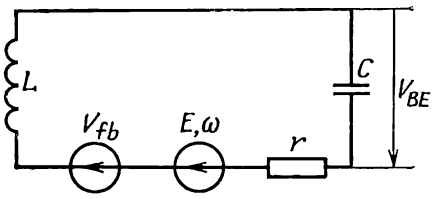


Fig. 9.33. Equivalent circuit of the tank shown in Fig. 9.32

written in a somewhat different form. Let us consider as an example a self-oscillator with the tank circuit inserted into the base-to-emitter circuit (Fig. 9.32). Figure 9.33 shows an equivalent circuit in which the feedback action is taken into account by introducing a voltage generator V_{fb}. When there is no external action, the amplitude V_{BE} is related to V_{fb} by the relation $V_{BE} = V_{fb}Q$.

Substituting this relation into formula (9.66) we get,

$$\frac{2\Delta\omega_{max}}{\omega_r} \approx E/V_{BE} \qquad (9.67)$$

In a similar way, we may show that when an external e.m.f. is applied to the collector tank circuit, the following relation holds:

$$\frac{2\Delta\omega_{max}}{\omega_r} \approx E/V_t \qquad (9.68)$$

where V_t is the amplitude of the voltage across the tank circuit of the free-running self-oscillator.

It is easy to notice that the absence of Q in formulas (9.67) and (9.68) is due to the fact that the external action is determined by the e.m.f. fed into the circuit in series, while the operating conditions of the free-running self-oscillator, by the voltage acting across the reactive element of the tuned circuit. If both voltages are defined identically—either across the reactive element or as the e.m.f. fed in series into the tuned circuit—the pulling range will be defined

by expression (9.66) regardless of the circuit configuration of the self-oscillator.

The relation between the oscillation frequency ω_{os} and the frequency ω of the driving e.m.f. is illustrated in Fig. 9.34. The dashed line in Fig. 9.34*a* corresponds to the case of the free-running self-oscillator, where $\omega_{os} = \omega_r = \text{const}$. Once the source of an external e.m.f. is introduced into the oscillator circuit, the frequency ω_{os}

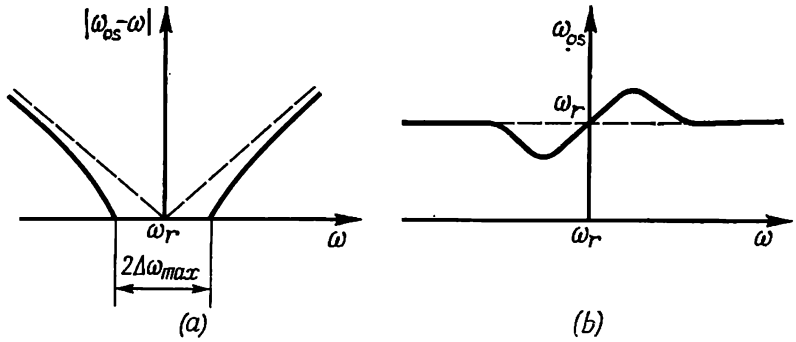


Fig. 9.34. Dependence of (*a*) the difference $|\omega_{os} - \omega|$ and (*b*) oscillation frequency ω_{os} on the frequency of the synchronizing source

is "pulled" to the frequency ω, and the solid line $|\omega_{os} - \omega|$ departs from the dashed line $|\omega_r - \omega|$. The segment of the abscissa axis, within which $|\omega_{os} - \omega| = 0$ and consequently, the oscillation frequency coincides with ω, is the pulling range.

The graph illustrating the dependence of the oscillation frequency ω_{os} on the frequency ω of the external e.m.f. is shown in Fig. 9.34*b* to a different frequency scale.

9.12. ANGLE MODULATION IN THE SELF-OSCILLATOR

In self-oscillators operating at frequencies not exceeding a few tens of megahertz, wide use is made of angle modulation methods based on direct variation of the resonance frequency of the tuned circuit of the oscillator by varying the capacitance or inductance of this circuit. Since the resonance frequency of the tuned circuit directly defines the oscillation frequency, by the term angle modulation in the self-oscillator we shall mean *frequency* modulation.

There are various methods of controlling the resonance frequency of a tuned circuit: electronic, electromagnetic, etc. The choice of a suitable method depends on the main modulation parameters — the relative frequency change $\Delta\omega/\omega_r$ and the rate of frequency change. The latter parameter is characterised by the spectrum of the modulating signal. In the case of slow modulation (low frequencies), wide use is made of methods whereby the inductance of the tank coil is varied by changing its magnetizing current, etc. If the message

spectrum includes comparatively high frequencies, it is necessary to employ lag-free methods for controlling the capacitance or inductance of the tank circuit.

One of the widely used methods for electronic control of the resonance frequency of a tuned circuit consists in that a varicap—a semiconductor diode (p-n junction) —whose capacitance depends on its reverse bias voltage is inserted into this circuit. A simplified circuit diagram of a self-oscillator with a varicap is shown in

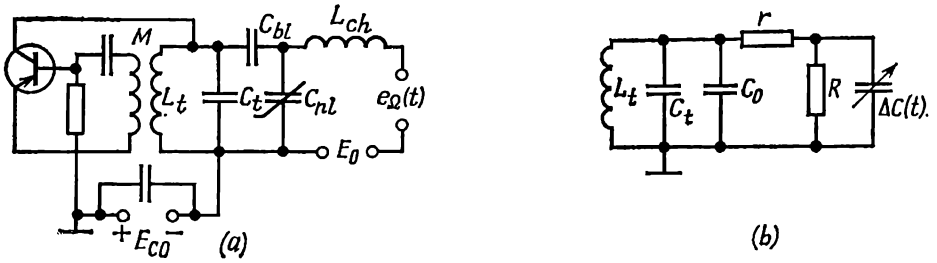


Fig. 9.35. (*a*) Self-oscillator with frequency modulation effected by means of a varicap and (*b*) the equivalent circuit of the tank

Fig. 9.35*a*. The blocking capacitor C_{bl} prevents the flow through the tank of the d-c current from the source of e.m.f. E_0 used for setting the operating point on the volt-farad characteristic of the varicap. The capacitor C_{bl} is also necessary to prevent the short-circuiting of the source of the modulating signal e_Ω (t) through the comparatively low inductance L_t of the tank circuit. The choke L_{ch} prevents the passage of the r-f current from the self-oscillator to the source of e.m.f. e_Ω (t).

In the equivalent circuit shown in Fig. 9.35*b*, C_0 stands for the average capacitance of the varicap in the absence of the modulating voltage e_Ω (t) and ΔC (t), for the capacitance variation proportional to the voltage e_Ω (t). The junction resistance is denoted by R and the volume resistance of the bulk of semiconductor, by r.

With the centre frequency ω_0 and frequency variation $\Delta\omega$ being specified, the required capacitance change ΔC can easily be found by means of the following obvious relations:

$$\omega_0 = 1/\sqrt{L_t C_{t0}}$$

$$\omega_0 + \Delta\omega = \frac{1}{\sqrt{L_t (C_{t0} + \Delta C)}} = \frac{1}{\sqrt{L_t C_{t0}}\sqrt{1 + \Delta C/C_{t0}}} = \frac{\omega_0}{\sqrt{1 + \Delta C/C_{t0}}}$$

where $C_{t0} = C_t + C_0$ is the average capacitance of the tank circuit.

Dividing the last expression by ω_0, we obtain

$$1 + \Delta\omega/\omega_0 = 1/\sqrt{1 + \Delta C/C_{t0}}$$

whence

$$\frac{\Delta C}{C_{t0}} = -\frac{2\Delta\omega/\omega_0 + (\Delta\omega/\omega_0)^2}{(1+\Delta\omega/\omega_0)^2} \tag{9.69}$$

In the general case, the required relative capacitance variation is related to the given relative frequency change by nonlinear relation (9.69). However, this relation has to be used only where frequency modulation is very deep. In many applications of frequency modulation, the relative frequency variation is very small. For example, when VHF band is used for transmitting music and speech, the value of $\Delta\omega/\omega_0$ does not exceed several fractions of a percent. In such cases, expression (9.69) can be simplified by neglecting $\Delta\omega/\omega_0$ in comparison with unity:

$$\Delta C/C_{t0} \approx -2\Delta\omega/\omega_0 \tag{9.70}$$

Thus, with small relative variations, $\Delta\omega$ and ΔC are related by linear relations, and to obtain linear frequency modulation, the capacitance should be varied in accordance with the modulating function $e_\Omega(t)$.

The disadvantage of the varicap as a frequency modulator is that the junction resistance R (Fig. 9.35*b*) depends on the amplitude of the external voltage. In the case of a comparatively deep frequency modulation requiring high amplitudes of the modulating voltage, this dependence causes a substantial change in the attenuation introduced into the tuned circuit of the self-oscillator and, in the final analysis, leads to parasitic amplitude modulation.

9.13. R-C OSCILLATORS

Oscillators with an *L-C* tank circuit are effective for generation of r-f oscillations, but for low (audio) frequencies they are inconvenient due to structural disadvantages (their tank circuits are cumbersome and difficult to retune). In this connection, *R-C* oscillators are widely used for generation of low-power low-frequency harmonic oscillations, especially in measuring instruments.

Two possible versions of an *R-C* oscillator are shown in Figs. 9.36 and 9.38. These oscillators differ from the *L-C* oscillator in that a resistance-coupled amplifier is used here instead of the amplifier with the tank circuit, and feedback is effected by means of a special two-port network made up of resistors and capacitors.

First, let us consider the version shown in Fig. 9.36. This circuit corresponds to an amplifier in which the phase of the voltage $\mathbf{V}_2$ is shifted through 180° with respect to $\mathbf{V}_1$ (e.g., a single-stage common-emitter transistor amplifier). To achieve oscillation at a specified frequency, it is required that the sum of the phase shifts in the closed feedback loop be equal to 2π and moreover, the ampli-

fication factor K_a be the inverse of K_{fb} [see formula (9.3)]. Therefore, the feedback two-port network enclosed in the dashed box in Fig. 9.36 must provide an additional phase shift of 180°.

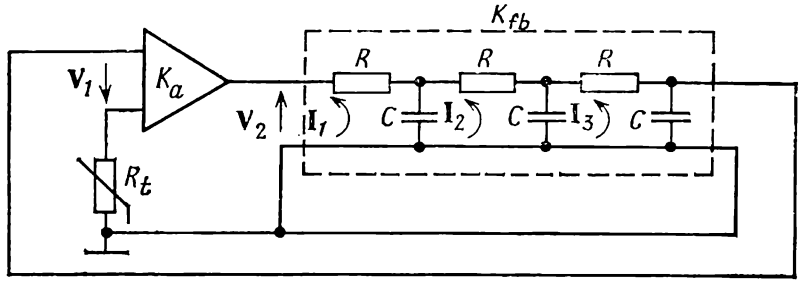


Fig. 9.36. Single-stage version of an *R-C* oscillator

The requirements for the elements of the two-port network can easily be revealed. Using the designations of Fig. 9.36, let us set up the following system of simultaneous equations:

$$\left.\begin{aligned} &(R+1/i\omega C)\,\mathbf{I}_1-(1/i\omega C)\,\mathbf{I}_2=\mathbf{V}_2\\ &-(1/i\omega C)\,\mathbf{I}_1+(R+2/i\omega C)\,\mathbf{I}_2-(1/i\omega C)\,\mathbf{I}_3=0\\ &-(1/i\omega C)\,\mathbf{I}_2+(R+2/i\omega C)\,\mathbf{I}_3=0 \end{aligned}\right\} \qquad (9.71)$$

Solving this system, we find

$$\mathbf{I}_3=\frac{-\mathbf{V}_2 i\omega C}{[5\,(\omega RC)^2-1]+i\,[(\omega RC)^3-6\omega RC]}$$

Since the voltage at the output of the feedback two-port network (read in the direction of the current $\mathbf{I}_3$) is equal to $\mathbf{I}_3/i\omega C$, the feedback factor is

$$\mathbf{K}_{fb}\,(i\omega)=\frac{\mathbf{I}_3/i\omega C}{\mathbf{V}_2}=-\frac{1}{[5\,(\omega RC)^2-1]+i\,[(\omega RC)^3-6\omega RC]}=K_{fb}e^{i\varphi_{fb}(\omega)} \qquad (9.72)$$

The argument and the modulus of the function $\mathbf{K}_{fb}\,(i\omega)$ are given in Fig. 9.37.

From expression (9.72) it follows that a 180-degree phase shift is obtained at a frequency satisfying the condition

$$\omega RC\,[(\omega RC)^2-6]=0$$

Consequently, oscillation is possible at the frequency

$$\omega_{os}=\sqrt{6}/RC \qquad (9.73)$$

There is a one-to-one relation between the modulus and the argument of the transfer function of the two-port under consideration. Substituting the value of ω_{os} thus found into expression (9.72),

we find the modulus

$$K_{fb}(\omega_{os}) = \frac{1}{[5(\omega_{os}RC)^2 - 1]} = \frac{1}{29}$$

Thus, the product RC defines the oscillation frequency and, at the same time, the feedback factor. This means that the gain of

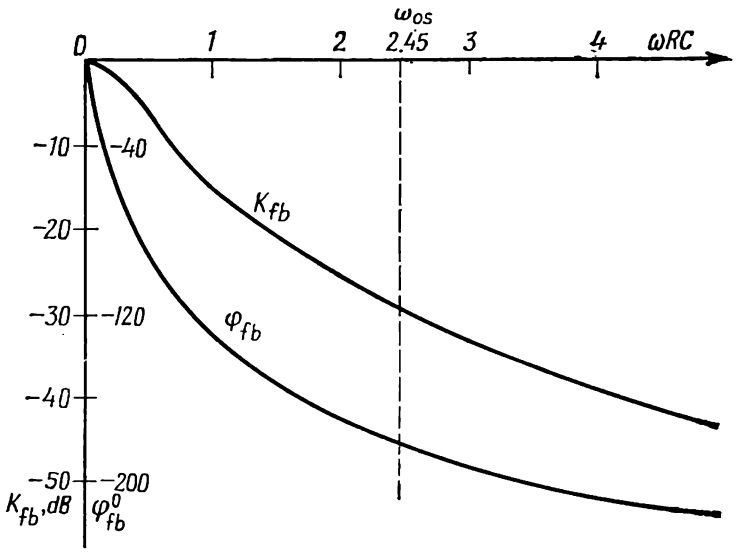


Fig. 9.37. Amplitude-frequency and phase-frequency characteristics of the two-port feedback network in the circuit shown in Fig. 9.36

the amplifier shown in Fig. 9.36 must be

$$K_a \geqslant 1/K_{fb} = 29$$

If the product RC is subdivided into multipliers, the choice of suitable values for R and C is much facilitated. It is only necessary to satisfy the condition $R \gg R_l$, where R_l is the load resistance of the amplifier, because only on this condition the amplification factor K_a is independent of R.

Stepwise variation of the oscillator frequency within a wide band is effected by means of a set of switch-selectable resistors and capacitors, while continuous tuning within each sub-band is carried out by means of variable capacitors.

Figure 9.38 shows another version of an R-C oscillator, in which the phase balance required for oscillation is ensured in the amplifier itself, e.g., by means of two stages, each of which shifts the phase through 180°. The auxiliary r_1-C_1-r_2-C_2 circuit is used for narrowing as far as possible the frequency band in which the phase balance is ensured.

To determine the relations between the elements r_1, C_1, r_2, and C_2, let us proceed from the transfer function of the feedback two-port network:

$$\mathbf{K}_{fb}(i\omega) = \frac{\dfrac{r_2/i\omega C_2}{r_2+1/i\omega C_2}}{(r_1+1/i\omega C_1)+\dfrac{r_2/i\omega C_2}{r_2+1/i\omega C_2}} = \frac{1}{1+r_1/r_2+C_2/C_1+i(\omega C_2 r_1-1/\omega C_1 r_2)} \tag{9.74}$$

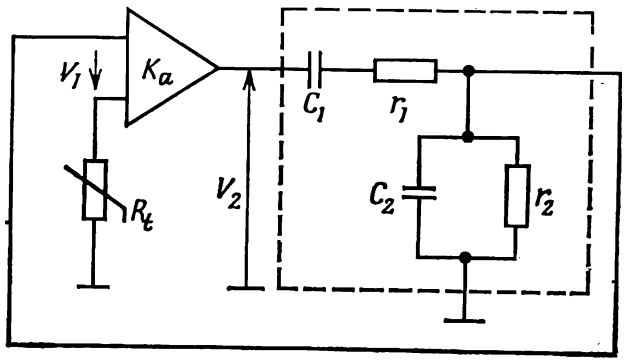


Fig. 9.38. Two-stage version of an *R-C* oscillator

In this case, the frequency at which oscillation is possible is defined by the condition

$$\omega_{os} C_2 r_1 - 1/\omega_{os} C_1 r_2 = 0$$

whence

$$\omega_{os} = 1/\sqrt{r_1 C_1 r_2 C_2} = 1/\sqrt{\tau_1 \tau_2} \tag{9.75}$$

where τ_1 and τ_2 are the time constants of the r_1-C_1 and r_2-C_2 circuits, respectively.

As a rule, identical resistors ($r_1 = r_2$) and capacitors ($C_1 = C_2$) are used. In this case, $\tau_1 = \tau_2 = \tau$, the oscillation frequency $\omega_{os} = 1/\tau$, and expression (9.74) assumes the form

$$\mathbf{K}_{fb}(i\omega) = \frac{1}{3+i(\omega/\omega_{os}-\omega_{os}/\omega)} = K_{fb}(\omega)\, e^{i\varphi_{fb}(\omega)}$$

where

$$K_{fb}(\omega) = \frac{1}{\sqrt{9+(\omega/\omega_{os}-\omega_{os}/\omega)^2}},$$

$$\varphi_{fb}(\omega) = -\arctan (1/3)(\omega/\omega_{os}-\omega_{os}/\omega)$$

The graphs of the modulus and the argument of the function $\mathbf{K}_{fb}(i\omega)$ for the selected parameters are shown in Fig. 9.39. It is essential that for any oscillation frequency, $K_{fb}(\omega_{os}) = 1/3 = \text{const}$. The independence of K_{fb} of the frequency ω_{os}, that pro-

vides for constant operating conditions of the oscillator within the entire frequency band, confirms the expediency of selecting $r_1 = r_2$, $C_1 = C_2$. Moreover, with $C_1 = C_2$, the construction of the gang capacitor used for smooth retuning of the oscillator is simplified.

The advantage of the circuit under consideration over that shown in Fig. 9.36 also consists in that the number of tunable elements here is smaller.

There are some other versions of *R-C* oscillators, but the above examples are sufficient for understanding the principle of self-oscillators with a periodic load and feedback circuits.

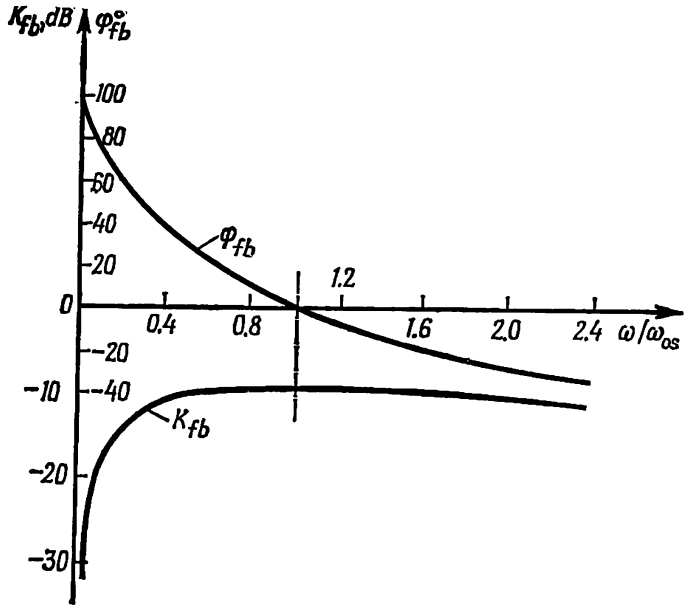


Fig. 9.39. Amplitude-frequency and phase-frequency characteristics of the two-port feedback network in the circuit shown in Fig. 9.38

The capacitors and resistors used in the feedback two-port network of any *R-C* oscillator must satisfy stringent requirements, for any instability of *R* or *C* due to temperature changes results in a change in the oscillation frequency. The capacitors must have high insulation resistance (low leakage), otherwise the shunting effect of the leakage in the range of very low frequencies will affect the phase relations in the two-port network.

Now let us analyze the mechanism of amplitude limitation in an *R-C* oscillator. This question is closely associated with that of the shape of the generated oscillations.

In the above-discussed *L-C* oscillators, the limitation is achieved owing to the fact that the average slope S_{av} decreases with increasing oscillation amplitude; steady-state conditions set in once the gain has reduced to $K_a = 1/K_{fb}$. However, in this case the ampli-

tude should not be allowed to go high, because this will cause distortion of the shape of the generated oscillations (due to the development of current harmonics). In contrast to *L-C* oscillators, *R-C* oscillators cannot provide for adequate suppression of higher harmonics. Thus, there is a contradiction between the requirement for undistorted shape of the oscillations (low amplitudes) and that for reliable limitation (high amplitudes). To eliminate this contradiction, *R-C* oscillators are usually provided with a time-lag nonlinear element in the form of a thermosensitive resistor (thermistor) whose resistance depends on the degree of its heating due to the current flowing through it. An ordinary incandescent lamp can be used as such a resistor.

The connection of a thermistor is illustrated in Figs. 9.36 and 9.38. It is implied that the feedback due to the introduction of R_t into the circuit is negative. For example, in a common-emitter transistor amplifier the resistor R_t is inserted into the emitter circuit.

The negative (current) feedback partially compensates for the positive feedback effected by means of the two-port network $\mathbf{K}_{fb}(i\omega)$. In fact, in the example under consideration the voltage produced across the resistor R_t by the a-c component of the collector current is directed from the emitter to the earthed line, while the positive feedback voltage, from the base to this line (see Fig. 9.36). Therefore, the resultant base-to-emitter potential difference is the difference between the second and the first voltage. The feedback factor K'_{fb}, understood as the ratio of the resultant base-to-emitter voltage to the collector-to-emitter voltage, depends on the value of R_t. When the oscillation amplitude and therefore, the current through the thermistor is increased, its resistance R_t increases and K'_{fb} decreases. On the contrary, when the oscillation amplitude is reduced, R_t decreases and K'_{fb} increases.

Thus, the amplitude limitation here is achieved owing to the decrease of K'_{fb} and not because the average slope S_{av} and the amplification factor K_a decrease with increasing oscillation amplitude. Steady-state conditions set in when $K'_{fb} = 1/K_a$. In this way, the oscillation amplitude is automatically controlled to stay at a definite level depending mainly on the nonlinearity of the thermistor characteristic. Since R_t changes comparatively slowly, because of the thermal inertia of the thermistor, within one oscillation period it remains practically constant. This implies that the variation of R_t does not cause nonlinear distortion and does not disturb the sinusoidal shape of the oscillations. The R_t in the circuit shown in Fig. 9.38 acts in a similar way. The *R-C* oscillator is a low-power device, and to obtain high power output, it is usually supplemented with one or two amplification stages.

Chapter 10

CIRCUITS WITH VARIABLE PARAMETERS

10.1. GENERAL

Electric circuits in which if only one parameter varies according to a given law are called *parametric circuits*. The variation (modulation) of the parameter is supposed to be effected electrically by means of a control oscillation.

Let us give some simple examples of electronic methods for varying the parameters of a circuit. Consider the dependence of the slope of the current-voltage characteristic $i(v)$ of an active element upon a control oscillation $e_c(t)$ superimposed on a constant voltage E_0 (Fig. 10.1). This dependence can be written in the form

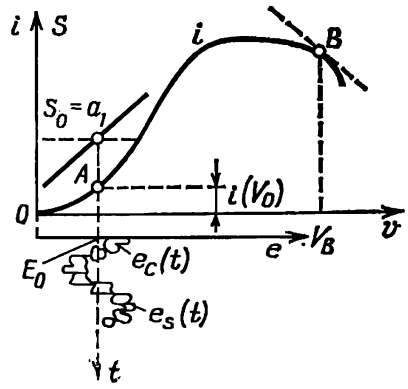


Fig. 10.1. Control of the slope of a characteristic curve

$$S(e_c) = \left(\frac{di}{dv}\right)_{v=E_0+e_c} \qquad (10.1)$$

Expression (10.1) defines the differential slope of the characteristic curve at the point $E_0 + e_c$. If within the range of e_c the characteristic can be approximated by the second-degree polynomial $i = i(E_0) + a_1(v - E_0) + a_2(v - E_0)^2$, expression (10.1) is reduced to the form

$$S(e_c) = a_1 + 2a_2 e_c = S_0 + 2a_2 e_c \qquad (10.2)$$

where $S_0 = a_1$ is the differential slope at the point A (Fig. 10.1). The dependence of the slope on the control voltage is shown in Fig. 10.1 in the form of an inclined straight line.

Let $e_c = E_c \cos \omega_c t$. In this case, the slope may be written in the form of a function of time

$$S(t) = a_1 + 2a_2 E_c \cos \omega_c t = a_1\left(1 + \frac{2a_2}{a_1} E_c \cos \omega_c t\right)$$

$$= S_0(1 + m \cos \omega_c t) \qquad (10.3)$$

where $m = 2a_2 F_c / a_1$ is the depth of "modulation" of the parameter S. By properly selecting E_0 and E_c, one can satisfy the condition $m < 1$.

With respect to a weak signal $e_s(t)$ superimposed on the control voltage $e_c(t)$, the device in question can be regarded as a linear circuit with a variable parameter $S(t)$ controlled according to (10.3). An essential feature of the differential slope (as well as of the differential resistance) is that it can take on a negative value. For this to occur, it is required that some section of the current-voltage characteristic (in the vicinity of the point B in Fig. 10.1) should have a negative inclination.

In a similar manner we can explain the principle of electronic capacitance control. Let a nonlinear capacitance be acted upon by two oscillations—a strong oscillation, which we shall call control oscillation, and a weak oscillation, which will be referred to as signal oscillation. Let us approximate the current-voltage characteristic shown in Fig. 8.2 by the second-degree polynomial

$$q = q_0 + b_1 (v - E_0) + 2b_2 (v - E_0)^2$$

Then, by analogy with (10.2), the differential capacitance can be defined by the expression

$$C(e_c) = \frac{dq}{dv}\Big|_{v=E_0+e_c} = b_1 + 2b_2 e_c$$

where $b_1 = C_0$ is the differential capacitance at the point $v = E_0$.

If the control voltage is a harmonic oscillation of the form $e_c = E_c \cos \omega_c t$, we may write

$$C(t) = C_0 (1 + m \cos \omega_c t) \tag{10.4}$$

where $m = 2b_2 E_c / b_1$ is the depth of modulation of the capacitance.

After such a transformation, one can consider the effect of *only one signal* $e_s(t)$ on the periodically varying capacitance $C(t)$, because the action of the control oscillation is taken account of by replacing the nonlinear capacitance by a linear parametric capacitance.

When using the barrier-layer capacitance of a p-n junction as the element being controlled, we can proceed from the expression

$$C(t) = C_0/(1 + m \cos \omega_c t) \tag{10.5}$$

where $m = E_c/2 (e_{con} + E_0)$, and e_{con} is the contact potential difference depending on the crystal, dopes, etc.

Similar expressions can also be set up for the current-controlled parametric inductance $L(t)$.

To establish the relationship between the charge of, current in, and voltage across the parametric capacitance, one should proceed from the obvious expressions

$$q(t) = C(t) v_C(t) \tag{10.6}$$

$$i(t) = \frac{dq}{dt} = C(t) \frac{dv_C}{dt} + v_C(t) \frac{dC}{dt} \tag{10.7}$$

$$v_C(t) = \frac{1}{C(t)} q(t) = \frac{1}{C(t)} \int i(t)\, dt \tag{10.8}$$

For the parametric inductance $L(t)$, the following relations between the magnetic linkage Φ, voltage v_L and current i hold true:

$$\Phi(t) = L(t)\, i(t) \tag{10.9}$$

$$v_L(t) = \frac{d\Phi}{dt} = L(t)\frac{di}{dt} + i(t)\frac{dL}{dt} \tag{10.10}$$

$$i(t) = \frac{1}{L(t)}\Phi(t) = \frac{1}{L(t)}\int v_L(t)\, dt \tag{10.11}$$

We should note the following principal difference between reactive and resistive elements: the *differential capacitance and inductance cannot be negative**. Physically, this is explained by the fact that increasing the voltage across a capacitor cannot decrease its charge, and increasing the current through an inductance coil cannot reduce its magnetic linkage. In other words, the energy stored in the electric field of the capacitor, or in the magnetic field of the coil, cannot be negative.

The elements with time-varying parameters $R(t)$, $C(t)$, and $L(t)$ will further be regarded as linear elements obeying the superposition principle. The terms "differential" resistance, "differential" capacitance or inductance, which are essential for characterizing the methods of varying the parameters but not for the analysis of the circuits composed of elements with such parameters, will not be used.

Circuits with variable parameters play an important part in radio engineering and electronics.

One can distinguish between two essentially different types of parameter variation:

(a) intentional, controlled variation for effecting various conversions of signals (modulation, frequency conversion, parametric amplification, etc.);

(b) uncontrollable variation due to various physical phenomena that occur when transmitting signals through free space (e.g., time-varying delay, variation of wave attenuation, phase shifts in the case of multibeam propagation of radio waves, fluctuation of the parameters of the radio channel, etc.).

The effect of variations of the second type, which are usually of statistical nature, will be discussed in Ch. 11. This chapter deals with phenomena occurring in the forced time variation of some parameter of a linear circuit (aperiodic or oscillatory). Mainly, it concerns harmonic variations.

---

* Here we mean ordinary elements. By using amplifiers with feedback one can simulate negative C and L.

10.2. TRANSMISSION OF OSCILLATIONS THROUGH LINEAR CIRCUITS WITH VARIABLE PARAMETERS. TRANSFER FUNCTION

In Chapters 6 and 7 we dealt with the transmission of various oscillations through linear networks with constant parameters. The relation between the input and the output signals in such circuits was established by means of either the transfer function $\mathbf{K}(i\omega)$ (the spectral method) or the impulse response $g(t)$ (the superposition integral method).

Now we shall consider a more general case where one or several elements of a linear two-port network are functions of time. It is evident that in such networks the relation between the output and

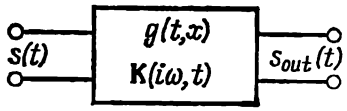


Fig. 10.2. Parametric two-port network

the input signal changes in the process of transmission. This means that the transfer function of the network depends not only on ω but also on time t. The impulse response also depends on two variables—the time interval $a = t - x$ between the moment x a unit impulse is applied to the input and the moment t the output signal appears (as in the case of a circuit with constant parameters) and the position of the interval a on the time axis. Therefore, the impulse response for a circuit with variable parameters should be written in the general form: $g(t, a)$ or $g(t, x)$.

We may say that $g(t, a)$ determines the response of the circuit at the moment t to the unit impulse applied to its input at the moment $x = t - a$. This impulse is written in the form of the delta function $\delta(t - x)$.

If an arbitrary signal $s(t)$ is applied to the input of a two-port network having an impulse response $g(t, x)$ (Fig. 10.2), then, according to the superposition principle and by analogy with (6.11), the output signal can be written as

$$s_{out}(t) = \int_{-\infty}^{\infty} s(x)\, g(t, x)\, dx \qquad (10.12)$$

Transforming to the variable a in accordance with the relation $x = t - a$, we obtain

$$s_{out}(t) = \int_{-\infty}^{\infty} s(t-a)\, g_1(t, a)\, da \qquad (10.13)$$

where $g_1(t, a) = g[t, (t - a)]$.

For a physically realizable circuit, $g_1(t, a) = 0$ for $a = t - x < 0$, i.e.., for $x > t$ (see Sec. 6.3).

Now let us introduce the transfer function $\mathbf{K}(i\omega, t)$ which is similar to the function $\mathbf{K}(i\omega)$ of a circuit with constant parameters, but accounts for the time variation of parameters. For this purpose, let us represent the function $s(t - a)$ in the form of a Fourier integral:

$$s(t-a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, e^{i\omega(t-a)}\, d\omega$$

where $\mathbf{S}(\omega)$ is the spectral density of the signal $s(t)$.

Then, expression (10.13) transforms into

$$s_{out}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, e^{i\omega t} \int_{-\infty}^{\infty} g_1(t, a)\, e^{-i\omega a}\, da\, d\omega$$

Denoting the internal integral by $\mathbf{K}(i\omega, t)$, we write the last expression as follows:

$$s_{out}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, \mathbf{K}(i\omega, t)\, e^{i\omega t}\, d\omega \qquad (10.14)$$

This expression coincides in form with analogous expression (6.2) for a circuit with constant parameters. From this it follows that the function $\mathbf{K}(i\omega, t)$ defined by the expression

$$\mathbf{K}(i\omega, t) = \int_{-\infty}^{\infty} g_1(t, a)\, e^{-i\omega a}\, da \qquad (10.15)$$

can be regarded as the *transfer function of a linear network with variable parameters.*

As in the case of a circuit with constant parameters, $\mathbf{K}(i\omega, t)$ is the Fourier transform of the impulse response $g_1(t, a)$. Since $g_1(t, a) = 0$ for $a < 0$, the lower limit of integration in (10.15) can be replaced by zero.

Along with expression (10.15), we can get another definition of the transfer function, in which the impulse response $g_1(t, a)$ is not present. For this, let us apply expression (10.14) to the case where the input signal is a harmonic oscillation $s(t) = \cos \omega_0 t$. Now we shall transform to the analytic signal $z(t) = e^{i\omega_0 t}$ corresponding to the given signal $s(t)$.

The spectral density of the analytic signal is $\mathbf{Z}(i\omega) = 2\pi \times \delta(\omega - \omega_0)$ [see expressions (2.98) and (3.90)]. Substituting $\mathbf{Z}(i\omega)$ for $\mathbf{S}(\omega)$ in formula (10.14), we get

$$z_{out}(t) = \int_{-\infty}^{\infty} \delta(\omega - \omega_0)\, \mathbf{K}(i\omega, t)\, e^{i\omega t}\, d\omega = \mathbf{K}(i\omega_0, t)\, e^{i\omega_0 t}$$

from which, omitting the zero subscript of ω, we obtain

$$\mathbf{K}(i\omega, t) = z_{out}(t)/e^{i\omega t} \tag{10.16}$$

where $z_{out}(t)$ is the analytic signal corresponding to the output signal $s_{out}(t)$. Thus,

$$s_{out}(t) = \operatorname{Re} z_{out}(t) = \operatorname{Re}[\mathbf{K}(i\omega, t)\, e^{i\omega t}] \tag{10.17}$$

If the transfer function $\mathbf{K}(i\omega)$ varies in time according to a periodic law with a frequency Ω, it can be represented in the form of a Fourier series:

$$\mathbf{K}(i\omega, t) = \mathbf{K}_0(i\omega) + \mathbf{K}_1(i\omega)\cos(\Omega t + \xi_1) + \mathbf{K}_2(i\omega)\cos(2\Omega t + \xi_2) + \ldots \tag{10.18}$$

where $\mathbf{K}_0(i\omega)$, $\mathbf{K}_1(i\omega)$, ... are coefficients, generally complex, which are independent of time and can be interpreted as the transfer functions of some two-port networks with *constant* parameters. The product $\mathbf{K}_n(i\omega)\cos(n\Omega t + \xi_n)$ can be regarded as the transfer function of two cascade-connected two-ports: one with the transfer function $\mathbf{K}_n(i\omega)$ independent of time and the other with the transfer function $\cos(n\Omega t + \xi_n)$ varying in time but independent of the frequency ω of the input signal.

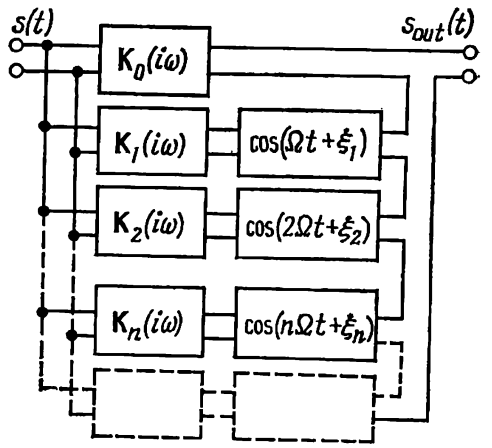


Fig. 10.3. Equivalent circuit of a linear network whose parameters vary periodically

In view of expression (10.18), any parametric system with periodically varying parameters can be represented in the form of an equivalent circuit as shown in Fig. 10.3. In accordance with (10.16) the output signal (complex) is

$$z_{out}(t) = \mathbf{K}(i\omega, t)\, e^{i\omega t} = \mathbf{K}_0(i\omega)\, e^{i\omega t} + \mathbf{K}_1(i\omega)\, e^{i\omega t}\cos(\Omega t + \xi_1)$$

$$+ \mathbf{K}_2(i\omega)\, e^{i\omega t}\cos(2\Omega t + \xi_2) + \ldots = K_0(\omega)\, e^{i(\omega t + \varphi_0)}$$

$$+ K_1(\omega)\, e^{i(\omega t + \varphi_1)}\cos(\Omega t + \xi_1) + K_2(\omega)\, e^{i(\omega t + \varphi_2)}\cos(2\Omega t + \xi_2) + \ldots \tag{10.19}$$

Here φ_0, φ_1, φ_2, ... are the phase characteristics of the two-port networks $\mathbf{K}_0(i\omega)$, $\mathbf{K}_1(i\omega)$, $\mathbf{K}_2(i\omega)$, ...

Transforming to the real output signal, we get

$$s_{out}(t) = \operatorname{Re} z_{out}(t) = K_0(\omega)\cos(\omega t + \varphi_0) + K_1(\omega)\cos(\omega t + \varphi_1)\cos(\Omega t + \xi_1) + K_2(\omega)\cos(\omega t + \varphi_2)\cos(2\Omega t + \xi_2) + \ldots = K_0(\omega)\cos(\omega t + \varphi_0) + \frac{1}{2}\sum_{n=1}^{\infty} K_n(\omega)\{\cos[(\omega + n\Omega)t + \varphi_n + \xi_n] + \cos[(\omega - n\Omega)t + \varphi_n - \xi_n]\} \quad (10.20)$$

This result is indicative of the following property of a circuit with variable parameters: *if the transfer function of the circuit varies according to any complex but periodic law with a fundamental frequency* Ω, *a harmonic input signal of frequency* ω *produces at the output of the circuit a spectrum including the frequencies* ω, $\omega \pm \Omega$, $\omega \pm 2\Omega$, etc.

If a complex signal is applied to the input of the circuit, all that has been said above applies to each frequency ω of the input spectrum. It is obvious that in a linear parametric circuit no individual components of the input spectrum interact (superposition principle), so that no frequencies of the form $n\omega_1 \pm m\omega_2$, where ω_1 and ω_2 are different frequencies of the input signal, exist at the output.

10.3. MODULATION AS A PARAMETRIC PROCESS

In Sec. 8.12 we considered the method of producing an AM oscillation, based on changing the amplitude current pulses in a *nonlinear* resonance amplifier. However, one should not draw a conclusion that modulation is a nonlinear process. In fact, the above method of modulation may be treated as the transmission of a carrier oscillation through a parametric two-port network whose transfer function varies according to the modulating voltage. (A change in the amplitude of current pulses and therefore, in the average slope S_{av} results in a change in the amplification factor of the circuit). Thus, disregarding the method for controlling circuit parameters, modulation should be considered a *parametric* process. This may be illustrated by the example of spectrum conversion in a simple parametric circuit, that was given in Sec. 1.5.

This argumentation fully applies to angle modulation. For example, let it be required to obtain an oscillation of the form $a(t) = A_0\cos[\omega_0 t + \theta(t) + \theta_0]$, where $\theta(t)$ is the phase angle modulated in accordance with a given law. Regarding $a(t)$ as an oscillation at the output of a linear two-port network whose input is acted upon by a carrier oscillation $e(t) = E_0\cos\omega_0 t$, let us find the transfer function of this two-port network. For this purpose, let us trans-

form from the given functions $e(t)$ and $a(t)$ to the complex oscillations $z_e(t) = E_0 e^{i\omega_0 t}$ and $z_a(t) = A_0 e^{i[\omega_0 t + \theta(t) + \theta_0]}$, respectively, and then use formula (10.16):

$$\mathbf{K}(i\omega_0, t) = \frac{A_0}{E_0} \frac{e^{i[\omega_0 t + \theta(t) + \theta_0]}}{e^{i\omega_0 t}} = K_0 e^{i[\theta(t+\theta_0)]} \qquad (10.21)$$

The two-port network with such a transfer function may be regarded as a delay line $\tau(t)$ satisfying the condition $\omega_0 \tau(t) = \theta(t) + \theta_0$.

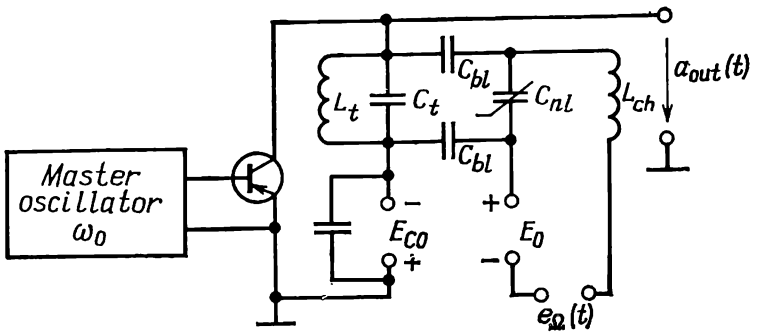


Fig. 10.4. Modulation of the phase of an oscillation by varying the resonance frequency of an amplifier

From this it follows that phase modulation requires a linear circuit with a delay $\tau(t)$ varying in time according to

$$\tau(t) = (1/\omega_0)[\theta(t) + \theta_0] \qquad (10.22)$$

Since the delay in a physical circuit cannot be negative, the term $\theta_0/\omega_0 = \tau_0$ having the sense of constant delay (in the absence of modulation) must not be smaller than $(1/\omega_0) \mid \theta(t) \mid_{\max}$.

The realization of a delay line allowing for electronic control of the quantity $\tau(t)$ is a complex problem. In the UHF band, it can be solved by using a travelling-wave tube in which the transit time can be varied to some extent by changing the potential on the corresponding electrodes. With a given constant excitation frequency ω_0 at the tube input, the change of the transit time by an amount $\Delta\tau$ is equivalent to the change of the phase of the output oscillation by an angle $\theta(t) = \omega_0 \Delta\tau$. Thus, it is possible to obtain fairly high values of $\theta(t)$ (tens of radians and more).

In the range of metric and longer waves, wide use is made of the methods based on changing the resonance frequency of the tuned circuit of an amplifier, the excitation frequency ω_0 being held constant. One of the possible versions of such a device is shown in Fig. 10.4. In this device, the resonance frequency of the tuned circuit is modulated by means of a varicap.

The relation between the variation ΔC of the capacitance of the tuned circuit and the variation $\Delta\omega$ of its resonance frequency was

established in Sec. 9.12. However, there is a fundamental difference between the devices shown in Figs. 9.35 and 10.4. A change in the resonance frequency of the tuned circuit in the self-oscillator (Fig. 9.35) is equivalent to a change in the oscillation frequency. But in the case of the amplifier (Fig. 10.4) which is excited from an independent source of carrier oscillation of frequency ω_0, a change in the resonance frequency ω_r of the tuned circuit *affects only the phase of the output oscillation.* The phase shift is easily found by the expression

$$\theta(t) = \arctan \frac{2\Delta\omega(t)}{\omega_r(t)} Q_{eq} \approx \arctan \frac{2\Delta\omega(t)}{\omega_0} Q_{eq} \qquad (10.23)$$

where $\Delta\omega(t) = \omega_r(t) - \omega_0$. The replacement of $\omega_r(t)$ by ω_0 in the denominator is usually permissible, for the relative variation of the resonance frequency ω_r in modulation is very small.

From expression (10.23) one can see the main disadvantage of this method of phase modulation, namely, the impossibility of obtaining high modulation indices. In fact, to provide for linear phase modulation (linear relation between $\theta(t)$ and $\Delta\omega(t)$ in the given case), the argument $2\Delta\omega(t) Q_{eq}/\omega_0$ must not exceed about 0.5 rad. This means that the phase variation amplitude $\theta_{max} = m$ is also limited to about 0.5 rad. At higher values of m, not only is the phase-modulating function distorted, but also a parasitic amplitude modulation arises, because in the case of substantial detuning the amplitude-frequency characteristic of the resonance amplifier is nonuniform. However, the method is advantageous in that the carrier frequency ω_0 can be made highly stable. The master oscillator operating at a fixed frequency can, for example, be stabilized by means of a quartz crystal.

10.4. IMPULSE RESPONSE OF A PARAMETRIC CIRCUIT

In order to determine the impulse response $g(t, x)$ directly by the given parameters of a circuit, without resorting to the transfer function $K(i\omega, t)$, it is necessary to use the differential equation of the circuit.

Let us consider a simple circuit described by the first-order equation

$$a_1(t)\frac{dy(t)}{dt} + a_0(t)\, y(t) = f(t) \qquad (10.24)$$

By definition, the impulse response is the response of the circuit to the unit impulse $\delta(t - x)$ applied to the input at the moment $t = x$ (see Sec. 10.2).

From this it follows that if the function $f(t)$ on the right-hand side of equation (10.24) is replaced by $\delta(t - x)$, then on the left-hand side $y(t)$ can be replaced by $g(t, x)$.

Thus, we come to the equation

$$a_1(t)\frac{dg(t, x)}{dt} + a_0(t)\, g(t, x) = \delta(t - x) \qquad (10.25)$$

Since the right-hand side of this equation is zero everywhere except for the point $t = x$, the function $g(t, x)$ can be found in the form of a solution of the homogeneous equation (with zero right-hand side)

$$a_1(t)\frac{dy}{dt} + a_0(t)\, y = 0 \qquad (10.26)$$

subject to the initial conditions stemming from equation (10.25) and also on condition that by the moment the impulse $\delta(t - x)$

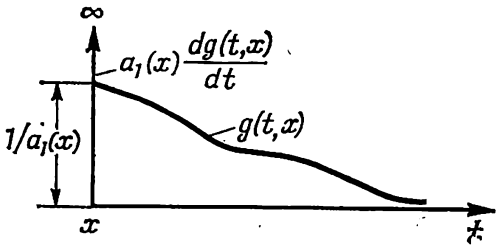


Fig. 10.5. Impulse response of the circuit described by equation (10.25)

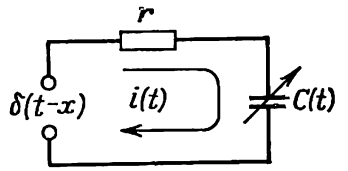


Fig. 10.6. Example of a simple parametric circuit

is applied there are no currents and voltages in the circuit ("empty" circuit).

In equation (10.26) the variables are separable:

$$\frac{dy}{y} + \frac{a_0(t)}{a_1(t)}dt = \frac{dy}{y} + P(t)\, dt = 0$$

whence

$$y = \varphi e^{-\int P(t)\, dt} \qquad (10.27)$$

where

$$P(t) = a_0(t)/a_1(t)$$

and

$$\varphi = y\,|_{t=x} = g(t, x)\,|_{t=x} \qquad (10.28)$$

is the value of the impulse response at the moment $t = x$.

To determine φ, let us return to the initial equation (10.25). From this equation it is clear that at the point $t = x$ the function $g(t)$ must step up by an amount $1/a_1(x)$ (Fig. 10.5). Only on this condition the first term in equation (10.25), i.e., $a_1(t)\, dg/dt$, can form the delta function $\delta(t - x)$.

Since for $t < x$ $g(t, x) = 0$, then at the moment $t = x$

$$g(t, x)_{t=x} = 1/a_1(x) \qquad (10.29)$$

Replacing in expression (10.27) the indefinite integral by a definite one with a variable upper limit, we obtain

$$g(t, x) = \varphi(x) \exp\left[-\int_x^t P(u)\, du\right] = \frac{1}{a_1(x)} \exp\left[-\int_x^t \frac{a_0(u)}{a_1(u)} du\right] \quad (10.30)$$

Here, the integration variable is denoted by u and not by t for the sake of clarity.

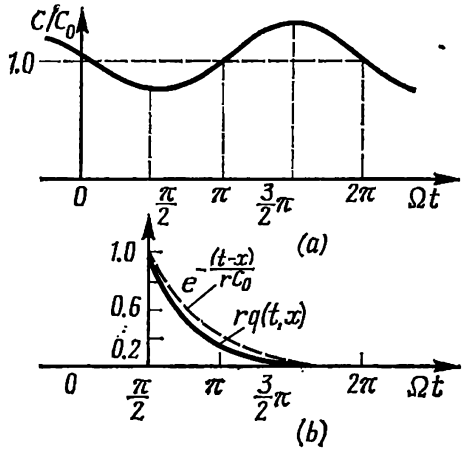


Fig. 10.7. (*a*) Law of capacitance modulation and (*b*) the impulse response of the circuit shown in Fig. 10.6

Let us apply expression (10.30) to a circuit (Fig. 10.6) consisting of a resistor and a capacitor connected in series, the resistance r of the resistor being constant and the capacitance $C(t)$ of the capacitor varying according to

$$C(t) = C_0/(1 + m \sin \Omega t) \quad (10.31)$$

In this case, by $\delta(t - x)$ is meant a unit e.m.f. impulse and the charge $q(t)$ of the capacitor is chosen to be the function $g(t, x)$ sought for.

Then, the circuit equation, in accordance with (10.25) and (10.31), can be written as

$$r\frac{dq}{dt} + \frac{1 + m \sin \Omega t}{C_0} q = \delta(t - x) \quad (10.32)$$

Substituting

$$a_1(t) = r, \quad a_0(t) = (1 + m \sin \Omega t)/C_0$$

into (10.30), we obtain

$$q(t, x) = \frac{1}{r} \exp\left(-\int_x^t \frac{1 + m \sin \Omega u}{rC_0} du\right)$$

$$= \frac{1}{r} \exp\left[-\frac{t - x}{rC_0} + \frac{m}{rC_0\Omega}(\cos \Omega t - \cos \Omega x)\right]$$

$$= \frac{1}{r} \exp\left\{-\frac{a}{rC_0} + \frac{m}{rC_0\Omega}[\cos \Omega t - \cos \Omega(t - a)]\right\} \quad (10.33)$$

Differentiating this expression with respect to t, one can find the current $i(t)$. At the moment $t = x$ (i.e., at $a = 0$), when $q(t, x)$ steps up by an amount $1/r$, the current is equal to $(1/r)\,\delta(t - x)$, while the voltage across the resistor is equal to $\delta(t - x)$. The voltage across the capacitor can be found by dividing expression (10.33) by $C(t)$.

From expression (10.33) one can see how the variation of the capacitance according to (10.31) affects the nature of the capacitor discharge: in addition to the term $-(t-x)/rC_0$ (as in the case of the constant capacitance C_0), in the exponent there appears the periodic term $(m/rC_0\Omega)$ [cos Ωt — cos $(t-a)$].

The law governing the variation of $C(t)/C_0$ is illustrated in Fig. 10.7*a*, while the graph of the function $rq(t, x)$ for $rC_0\Omega = 1$ and $m = 0.25$ is shown in Fig. 10.7*b*. In Fig. 10.7*b*, the dimensionless time Ωt is read from the moment $\Omega x = \pi/2$ at which $C(t)/C_0$ passes

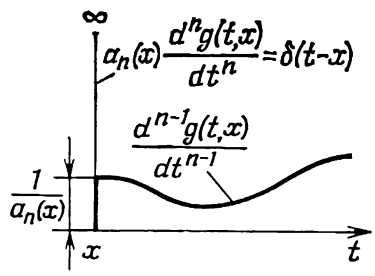


Fig. 10.8. Impulse response of the circuit described by equation (10.35)

through a minimum. The dashed line shows the graph of the function $e^{-(t-x)/rC_0}$ corresponding to the impulse response of the circuit with the constant capacitance C_0 $(m = 0)$.

Now let us consider the general case of a circuit described by the nth-order equation

$$a_n(t)\,y^n(t)+a_{n-1}(t)\,y^{n-1}(t)+\ldots+a_0(t)\,y(t)=f(t) \qquad (10.34)$$

n which all the coefficients a_n, a_{n-1}, . . ., a_0 may be functions of t (but not of y).

As in the preceding case, let us equate $f(t)$ to the delta function $\delta(t-x)$, which allows equation (10.34) to be rewritten in the form

$$a_n(t)\frac{d^n g}{dt^n}+a_{n-1}(t)\frac{d^{n-1}g}{dt^{n-1}}+\ldots+a_0(t)\,g=\delta(t-x) \qquad (10.35)$$

The impulse response is found in the form of a solution of the homogeneous equation

$$a_n(t)\frac{d^n y}{dt^n}+a_{n-1}(t)\frac{d^{n-1}y}{dt^{n-1}}+\ldots+a_0(t)\,y=0 \qquad (10.36)$$

The general solution of the equation is a sum of n linearly independent solutions

$$g(t, x)=\sum_{i=1}^{n}\varphi_i(x)\,y_i(t) \qquad (10.37)$$

The functions $\varphi_i(x)$ can be found by using the initial conditions stemming from equation (10.34).

In contrast to equation (10.25), in this case it is necessary at the point $t = x$ to equate to zero all the derivatives of the function $g(t, x)$ of the order not higher than $n-2$. The derivative of the order $n-1$ must step up by an amount $1/a_n(x)$ at the same point (Fig. 10.8).

Subject to these initial conditions, expression (10.37) forms a system of equations (after differentiating with respect to t and substituting $t = x$):

$$\left.\begin{aligned}
&\varphi_1(x)\,y_1(x)+\varphi_2(x)\,y_2(x)+\ldots+\varphi_n(x)\,y_n(x)=0\\
&\varphi_1(x)\frac{dy_1(t)}{dt}\Big|_{t=x}+\varphi_2(x)\frac{dy_2(t)}{dt}\Big|_{t=x}+\ldots+\varphi_n(x)\frac{dy_n(t)}{dt}\Big|_{t=x}=0\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\varphi_1(x)\frac{d^{n-2}y_1(t)}{dt^{n-2}}\Big|_{t=x}+\varphi_2(x)\frac{d^{n-2}y_2(t)}{dt^{n-2}}\Big|_{t=x}+\ldots\\
&\qquad+\varphi_n(x)\frac{d^{n-2}y_n(t)}{dt^{n-2}}\Big|_{t=x}=0\\
&\varphi_1(x)\frac{d^{n-1}y_1(t)}{dt^{n-1}}\Big|_{t=x}+\varphi_2(x)\frac{d^{n-1}y_2(t)}{dt^{n-1}}\Big|_{t=x}+\ldots\\
&\qquad+\varphi_n(x)\frac{d^{n-1}y_n(t)}{dt^{n-1}}\Big|_{t=x}=\frac{1}{a_n(x)}
\end{aligned}\right\}\qquad(10.38)$$

Since the partial solutions $y_i(t)$ and their derivatives at the point $t = x$ are supposed to be known, system (10.38) containing n equations makes it possible to find all the functions $\varphi_i(x)$. System (10.38) has a corresponding determinant

$$W(x)=\begin{vmatrix} y_1(x) & y_2(x) & \ldots & y_n(x)\\ y_1'(x) & y_2'(x) & \ldots & y_n'(x)\\ \ldots & \ldots & \ldots & \ldots\\ y_1^{(n-1)}(x) & y_2^{(n-1)}(x) & \ldots & y_n^{(n-1)}(x)\end{vmatrix}$$

known as the *Wronskian determinant* (*Wronskian*).

By applying Cramer's rule, we obtain the following general expression for $\varphi_i(x)$:

$$\varphi_i(x)=\frac{(-1)^{n+i}M_{ni}(x)}{a_n(x)\,W(x)}\qquad(10.39)$$

where $M_{ni}(x)$ is the minor obtained from the determinant $W(x)$ by erasing the nth row (corresponding to the derivatives of the order $n-1$) and the ith column which intersect at the element $y_i^{(n-1)}(x)$.

Substituting (10.39) into general solution (10.37), we obtain

$$g(t,x)=\frac{1}{a_n(x)\,W(x)}\left[(-1)^{n+1}M_{n1}y_1(t)\right.$$
$$\left.+(-1)^{n+2}M_{n2}y_2(t)+\ldots+(-1)^{n+n}M_{nn}y_n(t)\right]\qquad(10.40)$$

The expression in square brackets can be regarded as the expansion of the corresponding determinant along the row $y_1(t)$, $y_2(t)$, ..., $y_n(t)$. Denoting this new determinant by d, we may write

$$d=A_{11}y_1(t)+A_{12}y_2(t)+\ldots+A_{1n}y_n(t)$$

from which follows the equality for cofactors A_{1i}

$$A_{1i}=(-1)^{n+i}M_{ni}\qquad(10.41)$$

On the other hand, in the determinant d the cofactors A_{1i} are related to the corresponding minors M_{1i}^d by the relation

$$A_{1i}=(-1)^{1+i}M_{1i}^d$$

Consequently, the minors of the new determinant are expressed through the minors of the Wronskian determinant by means of the following formula:

$$M_{ii}^{d}=\frac{(-1)^{n+i}M_{ni}}{(-1)^{1+i}}=(-1)^{n-1}M_{ni}$$

But, as has been mentioned above, the minor M_{ni} is obtained from the Wronskian determinant by erasing the ith column and the row corresponding to the derivatives of the order $n-1$. Consequently, the sought for determinant d must be written as

$$d=(-1)^{n-1}\begin{vmatrix} y_1(t) & y_2(t) & \dots & y_n(t) \\ y_1(x) & y_2(x) & \dots & y_n(x) \\ \dots & \dots & \dots & \dots \\ y_1^{(n-2)}(x) & y_2^{(n-2)}(x) & \dots & y_n^{(n-2)}(x) \end{vmatrix}$$

Thus, finally,

$$g(t, x)=\frac{(-1)^{n-1}}{a_n(x)W(x)}\begin{vmatrix} y_1(t) & y_2(t) & \dots & y_n(t) \\ y_1(x) & y_2(x) & \dots & y_n(x) \\ \dots & \dots & \dots & \dots \\ y_1^{(n-2)}(x) & y_2^{(n-2)}(x) & \dots & y_n^{(n-2)}(x) \end{vmatrix} \qquad (10.42)$$

The function $g(t, x)$ determined by expression (10.42) is none other than the *unilateral Green's function* of the linear differential operator

$$\mathrm{L}=a_n(t)p^n+a_{n-1}(t)p^{n-1}+\dots+a_0(t)$$

corresponding to equation (10.34). In the theory of linear nonhomogeneous equations, the Green's function is used to represent the solution of equation (10.34) in the form

$$y(t)=\int_{-\infty}^{t} g(t, x)f(x)\,dx \qquad (10.43)$$

subject to the initial conditions $y^{(k)}(0)=0$, $k=0, 1, \dots, (n-1)$.

This expression coincides with (10.13).

10.5. ENERGY RELATIONS IN A CIRCUIT WITH A NONLINEAR REACTIVE ELEMENT IN THE CASE OF BIHARMONIC EXCITATION

In Sec. 10.1 it was noted that the electronic method for obtaining a variable capacitance (linear) is based on the use of a nonlinear capacitance acted upon by a control oscillation.

A specific feature of the biharmonic (signal and control oscillation) excitation of a nonlinear reactive element is the energy exchange between the sources of the individual oscillations. The energy consumed in a parametric circuit is supplied ("pumped") by the source of the control oscillation. In this connection, it is often called pumping oscillator, while the control oscillation is called pumping voltage. Therefore, later in the text, the control oscillation $e_c(t)=E_c\cos(\omega_c t+\theta_c)$ will often be written in the form $e_p(t)=E_p\cos(\omega_p t+\theta_p)$.

Let us consider the action on a nonlinear capacitor C_{nl} (a varicap, Fig. 10.9) of two harmonic oscillations:

$$e(t) = e_1(t) + e_p(t) = E_1 \cos(\omega_1 t + \theta_1) + E_p \cos(\omega_p t + \theta_p) \quad (10.44)$$

As in Sec. 10.1, let us approximate the volt-coulomb characteristic of the varicap by the second-degree polynomial

$$q = q_0 + b_1 e + b_2 e^2 \quad (10.45)$$

where $b_1 = C_0$ is defined by expression (8.4), while

$$b_2 = \frac{1}{2!}\left(\frac{d^2q}{de^2}\right)_{e=E_0} = \frac{1}{2!}\left(\frac{dC}{de}\right)_{e=E_0} \quad (10.46)$$

Applying the first expression of (8.3) to series (10.45), we find the current flowing through the nonlinear capacitor:

$$i(t) = \frac{dq}{de}\frac{de}{dt} = (b_1 + 2b_2 e)\frac{de}{dt} = b_1\frac{de}{dt} + 2b_2 e\frac{de}{dt} \quad (10.47)$$

Substituting into this expression the excitation (10.44), we get (after simple trigonometric transformations) the final expression

$$\begin{aligned} i(t) = -\{&b_1\omega_1 E_1 \sin(\omega_1 t + \theta_1) + b_1\omega_p E_p \sin(\omega_p t + \theta_p) \\ &- b_2\omega_1 E_1^2 \sin 2(\omega_1 t + \theta_1) + b_2\omega_p E_p^2 \sin 2(\omega_p t + \theta_p) \\ &+ b_2(\omega_p + \omega_1) E_1 E_p \sin[(\omega_p + \omega_1)t + (\theta_p + \theta_1)] \\ &- b_2(\omega_p - \omega_1) E_1 E_p \sin[(\omega_p - \omega_1)t + (\theta_p - \theta_1)]\} \end{aligned} \quad (10.48)$$

The first two terms in the expression thus obtained correspond to currents of frequencies ω_1 and ω_p, that would occur in the circuit

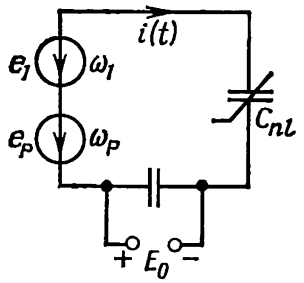


Fig. 10.9. Biharmonic action on a nonlinear capacitor

with a linear capacitance equal to b_1; the remaining terms—harmonics with frequencies $2\omega_1$ and $2\omega_p$ and combination oscillations with frequencies $\omega_p + \omega_1$ and $\omega_p - \omega_1$ are the products of interaction of the two harmonic oscillations in the quadratic nonlinear element.

It should be noted that the spectrum of the current through a nonlinear capacitor differs from that of the current through a nonlinear resistive element only in that there is no d-c component [see (8.28) through (8.30)].

Let us consider the energy relations. The first two currents (those with coefficients b_1), that are shifted in phase by 90° relative to the

respective e.m.f.'s $e_1(t)$ and $e_p(t)$, do not cause consumption of energy (as in an ordinary linear loss-free capacitor). The current of frequency $\omega_p + \omega_1$ can also be disregarded in energy calculations, since the average power produced either by the oscillator of frequency ω_1 or by that of frequency ω_p is zero when a current of frequency $\omega_p + \omega_1$ flows through it.

On the other hand, the load on the oscillators due to the currents of frequencies $2\omega_1$, $2\omega_p$, and $\omega_p - \omega_1$ depends on the relations between the frequencies ω_p and ω_1 and between the phases θ_1 and θ_p. If the frequency ω_p is a multiple of ω_1 and its number is the same as the degree of the approximating polynomial (i.e., the order of nonlinearity), these currents, depending on the relation between the phases θ_1 and θ_p, may be in phase (or in antiphase) with the electromotive forces $e_1(t)$ or $e_p(t)$.

In particular, in the given case of quadratic nonlinearity, if the condition $\omega_p = 2\omega_1$ is satisfied, the combination frequency $\omega_p - \omega_1$ coincides with ω_1 (correspondingly, if $\omega_1 = 2\omega_p$, $|\omega_p - \omega_1| = \omega_1/2 = \omega_p$). In this case, the currents of frequencies $2\omega_p$ and $\omega_p + \omega_1$ may be ignored and the currents of frequencies ω_p and ω_1 represented in the form

$$i_{\omega_p}(t) = -b_2\omega_1 E_1^2 \sin 2(\omega_1 t + \theta_1) = -b_2\omega_1 E_1^2 \sin(\omega_p t + 2\theta_1)$$
$$= b_2\omega_1 E_1^2 \cos(\omega_p t + 2\theta_1 + \pi/2)$$

$$i\omega_1(t) = -b_2(\omega_p - \omega_1) E_1 E_p \sin[(\omega_p - \omega_1) t + (\theta_p - \theta_1)]$$
$$= -b_2\omega_1 E_1 E_p \sin[\omega_1 t + (\theta_p - \theta_1)]$$
$$= b_2\omega_1 E_1 E_p \cos[\omega_1 t + (\theta_p - \theta_1) + \pi/2]$$

Since the current $i_{\omega_p}(t)$ is shifted in phase by an angle $\theta_p - (2\theta_1 + \pi/2) = \theta_p - 2\theta_1 - \pi/2$ with respect to $e_p(t) = E_p \cos(\omega_p t + \theta_p)$, the average power delivered by the source of the e.m.f. $e_p(t)$ is

$$P_{\omega_p} = \frac{1}{2} b_2\omega_1 E_1^2 E_p \cos(\theta_p - 2\theta_1 - \pi/2)$$
$$= \frac{1}{2} b_2\omega_1 E_1^2 E_p \sin(\theta_p - 2\theta_1) \qquad (10.49)$$

Correspondingly, the average power delivered by the oscillator of frequency ω_1 in the case of the phase shift $\theta_1 - [(\theta_p - \theta_1) + \pi/2] = 2\theta_1 - \theta_p - \pi/2$ is

$$P_{\omega_1} = \frac{1}{2} b_2\omega_1 E_1^2 E_p \cos(2\theta_1 - \theta_p - \pi/2)$$
$$= \frac{1}{2} b_2\omega_1 E_1^2 E_p \sin(2\theta_1 - \theta_p) \qquad (10.50)$$

Note that the condition

$$P_{\omega_1} + P_{\omega_p} = 0$$

is satisfied no matter what the phases θ_1 and θ_p.

Let us impose an additional condition: $\theta_p - 2\theta_1 = \pi/2$. In this case, P_{ω_p} is positive, while P_{ω_1} is negative. This means that the oscillator of frequency ω_1 does not deliver power but, on the contrary, consumes it from the oscillator of frequency ω_p.

Now let us set $\theta_p - 2\theta_1 = -\pi/2$. Then, $P_{\omega_p} < 0$ and $P_{\omega_1} > 0$. In this case, the power source is the oscillator of frequency ω_1 and the power consumer is the oscillator of frequency ω_p.

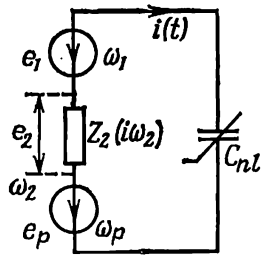


Fig. 10.10. Biharmonic action on a network with a nonlinear capacitor and a parallel resonant circuit tuned to a combination frequency

Now let us consider energy relations in a circuit containing, in addition to the nonlinear capacitor C_{nl}, a linear dissipative element — a parallel resonant circuit tuned to a frequency close to $\omega_2 = \omega_p - \omega_1$. In Figure 10.10 this element is denoted by $Z_2(i\omega_2)$ (impedance).

The nonlinear capacitor is acted upon by two electromotive forces — the signal $e_1(t) = E_1\cos(\omega_1 t + \theta_1)$ and the pump voltage $e_p(t) = E_p\cos(\omega_p t + \theta_p)$ — supplied from two independent sources and, in addition, by a voltage of combination frequency $\omega_2 = \omega_p - \omega_1$, which is a product of interaction of $e_1(t)$ and $e_p(t)$ in the nonlinear capacitor. Let us write this voltage in the form $e_2(t) = E_2\cos(\omega_2 t + \theta_2)$, bearing in mind that the amplitude E_2 and phase θ_2, that depend not only on the oscillations $e_1(t)$ and $e_p(t)$, but also on the impedance $Z_2(i\omega_2)$, have yet to be defined. In the text below, we shall proceed from the condition that the impedance $Z_2(i\omega_2)$ is negligibly low for the frequencies ω_1 and ω_p. Substituting

$$e(t) = E_1\cos(\omega_1 t + \theta_1) + E_2\cos(\omega_2 t + \theta_2) + E_p\cos(\omega_p t + \theta_p)$$

into expression (10.47), we obtain, after simple manipulations, a formula similar to (10.48) (the frequencies other than $\omega_p - \omega_2 =$

$= \omega_1$, $\omega_p - \omega_1 = \omega_2$, and $\omega_1 + \omega_2 = \omega_p$ are neglected):

$$\begin{aligned} i(t) = &-b_1[\omega_1 E_1 \sin(\omega_1 t + \theta_1) + \omega_2 E_2 \sin(\omega_2 t + \theta_2) \\ &+ \omega_p E_p \sin(\omega_p t + \theta_p)] - b_2\{\omega_1 E_2 E_p \sin[\omega_1 t + (\theta_p - \theta_2)] \\ &+ \omega_2 E_1 E_p \sin[\omega_2 t + (\theta_p - \theta_1)] \\ &- \omega_p E_1 E_2 \sin[\omega_p t + (\theta_1 + \theta_2)]\} \end{aligned} \tag{10.51}$$

When determining the energy balance, the first three terms (with coefficients b_1) can be ignored [see the comments to formula (10.48)]. The currents of frequencies ω_1, ω_2, and ω_p, that are due to the nonlinearity of the volt-coulomb characteristic, are defined by the expressions

$$\left.\begin{aligned} i_{\omega_1}(t) &= -b_2\omega_1 E_2 E_p \sin[\omega_1 t + (\theta_p - \theta_2] \\ &= I_{\omega_1} \cos[\omega_1 t + (\theta_p - \theta_2) + \pi/2] \\ i_{\omega_2}(t) &= -b_2\omega_2 E_1 E_p \sin[\omega_2 t + (\theta_p - \theta_1)] \\ &= I_{\omega_2} \cos[\omega_2 t + (\theta_p - \theta_1) + \pi/2] \\ i_{\omega_p}(t) &= +b_2\omega_p E_1 E_2 \sin[\omega_p t + (\theta_1 + \theta_2)] \\ &= I_{\omega_p} \cos[\omega_p t + (\theta_1 + \theta_2) + \pi/2] \end{aligned}\right\} \tag{10.52}$$

Here $I_{\omega_1} = b_2\omega_1 E_2 E_p$, $I_{\omega_2} = b_2\omega_2 E_1 E_p$, and $I_{\omega_p} = b_2\omega_p E_1 E_2$ are the amplitudes of the currents $i_{\omega_1}(t)$, $i_{\omega_2}(t)$, and $i_{\omega_p}(t)$.

Taking into account that in the initial expressions $e_2(t) = E_2 \cos(\omega_2 t + \theta_2)$ has the sense of the e.m.f. compensating for the voltage drop across the impedance $Z_2(i\omega_2)$ due to the current $i_{\omega_2}(t)$ flowing through it and using the second equation of (10.52), we can set up the following expression

$$\begin{aligned} e_2(t) = E_2 \cos(\omega_2 t + \theta_2) = &-I_{\omega_2} Z_2(\omega_2) \cos[\omega_2 t + (\theta_p - \theta_1) \\ &+ \pi/2 + \varphi_Z] = I_{\omega_2} Z_2(\omega_2) \cos[\omega_2 t + (\theta_p - \theta_1) - \pi/2 + \varphi_Z] \end{aligned}$$

where φ_Z stands for the argument of the impedance $Z_2(i\omega_2)$. From this it follows that

$$E_2 = I_{\omega_2} Z_2(\omega_2), \quad \theta_2 = \theta_p - \theta_1 - \pi/2 + \varphi_Z$$

Using the last expression, formulas (10.52) can be reduced to the following form:

$$\left.\begin{aligned} i_{\omega_1}(t) &= I_{\omega_1} \cos[\omega_1 t + (\theta_1 - \varphi_Z + \pi/2) + \pi/2] \\ &= -I_{\omega_1} \cos(\omega_1 t + \theta_1 - \varphi_Z) \\ i_{\omega_2}(t) &= I_{\omega_2} \cos[\omega_2 t + (\theta_2 - \varphi_Z + \pi/2) + \pi/2] \\ &= -I_{\omega_2} \cos(\omega_2 t + \theta_2 - \varphi_Z) \\ i_{\omega_p}(t) &= I_{\omega_p} \cos[\omega_p t + (\theta_p + \varphi_Z + \pi/2) - \pi/2)] \\ &= +I_{\omega_p} \cos(\omega_p t + \theta_p + \varphi_Z) \end{aligned}\right\} \tag{10.53}$$

Let us set up expressions for the power dissipated in the nonlinear capacitor at frequencies ω_1, ω_2, and ω_p.

$$\left.\begin{aligned} P_{\omega_1} &= -\frac{1}{2} I_{\omega_1} E_1 \cos \varphi_Z = -\frac{1}{2} b_2 \omega_1 E_1 E_2 E_p \cos \varphi_Z \\ P_{\omega_2} &= -\frac{1}{2} I_{\omega_2} E_2 \cos \varphi_Z = -\frac{1}{2} b_2 \omega_2 E_1 E_2 E_p \cos \varphi_Z \\ P_{\omega_p} &= \frac{1}{2} I_{\omega_p} E_p \cos \varphi_Z = \frac{1}{2} b_2 \omega_p E_1 E_2 E_p \cos \varphi_Z \end{aligned}\right\} \quad (10.54)$$

The negative values of P_{ω_1} and P_{ω_2} mean that the corresponding sources of the e.m.f.'s $e_1(t)$ and $e_2(t)$ consume energy instead of delivering it. On the other hand, the positive value of P_{ω_p} indicates to the fact that at the given frequencies ($\omega_p > \omega_2$, $\omega_p > \omega_1$) the source of the e.m.f. $e_p(t)$ delivers energy to the external circuit.

The total power dissipated in the nonlinear reactive element is

$$P_{\omega_1} + P_{\omega_2} + P_{\omega_p} = \frac{1}{2} b_2 (\omega_p - \omega_1 - \omega_2) E_1 E_2 E_p \cos \varphi_Z = 0$$

since $\omega_p = \omega_1 + \omega_2$. This result is fully consistent with the premise that there are no losses in the capacitor.

From expressions (10.54) we obtain the following proportions

$$P_{\omega_1}/\omega_1 = P_{\omega_2}/\omega_2 = -P_{\omega_p}/\omega_p \quad (10.55)$$

These proportions are a particular case of the Manly-Raw general theorem of energy relations in the oscillation spectrum in a circuit containing a nonlinear reactive element. This theorem is written in the form

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{m\omega_1 + n\omega_0} = 0, \quad \sum_{n=0}^{\infty} \sum_{m=-\infty}^{\infty} \frac{nP_{m,n}}{m\omega_1 + n\omega_0} = 0 \quad (10.56)$$

where ω_1 and ω_0 are the frequencies of the oscillators exciting the system and $P_{m,n}$ is the power of the oscillation of frequency $m\omega_1 + n\omega_0$, the integers m and n determining the order of the combination oscillation. In the general case, it is assumed that in a circuit with a nonlinear reactance there are admittances for any combination frequencies.

Expressions (10.56) can be extended to cover any reactances — capacitive and inductive — provided there is no hysteresis.

The first equation of (10.56), in which m takes on only positive values, establishes the relation between the power $P_{m,n}$ of the combination oscillation and the frequency ω_1 of the oscillator. Correspondingly, the second equation, in which $n \geqslant 0$, establishes the relation between the power of the combination oscillations and the frequency ω_0 of the second oscillator.

Let us illustrate the application of expressions (10.56), taking as an example the previously studied circuit (Fig. 10.10) excited

by two oscillators of frequencies ω_1 and $\omega_0 = \omega_p$. Besides these frequencies, a combination oscillation of difference frequency $\omega_2 = \omega_p - \omega_1$ is produced in the passive element $Z_2(\omega_2)$.

In accordance with the notation of expressions (10.56), the frequency ω_1 should be regarded as the value of the denominator $m\omega_1 + n\omega_0$ at $m = 1$ and $n = 0$, and the power at this frequency $P_{\omega_1} = P_{1,0}$. The summation indices $m = 0$ and $n = 1$ and the power $P_{\omega_0} = P_{0,1}$ correspond to the frequency ω_0. Finally, the summation indices $m = -1$ and $n = 1$ and the power $P_{\omega_2} = P_{\omega_0-\omega_1} = P_{-1,1}$ correspond to the frequency $\omega_2 = \omega_0 - \omega_1$.

Then, the internal sum in the first equation of (10.56) yields

$$\sum_{n=-1}^{1} \frac{mP_{m,n}}{m\omega_1+n\omega_0} = \frac{mP_{m,-1}}{m\omega_1-\omega_0} + \frac{mP_{m,0}}{m\omega_1+0\times\omega_0} + \frac{mP_{m,1}}{m\omega_1+\omega_0}$$

$$= \frac{mP_{m,-1}}{m\omega_1-\omega_0} + \frac{P_{m,0}}{\omega_1} + \frac{mP_{m,1}}{m\omega_1+\omega_0}$$

Summing the obtained expression over m, we get the first equation of (10.56):

$$\sum_{m=0}^{1}\left(\frac{mP_{m,-1}}{m\omega_1-\omega_0} + \frac{P_{m,0}}{\omega_1} + \frac{mP_{m,1}}{m\omega_1+\omega_0}\right)$$

$$= \left(0 + \frac{P_{0,0}}{\omega_1} + 0\right) + \left(\frac{P_{1,-1}}{\omega_1-\omega_0} + \frac{P_{1,0}}{\omega_1} + \frac{P_{1,1}}{\omega_1+\omega_0}\right) = \frac{P_{1,-1}}{\omega_1-\omega_0} + \frac{P_{1,0}}{\omega_1} = 0$$

(The terms including $P_{0,0}$ and $P_{1,1}$ are omitted). Thus,

$$P_{\omega_0-\omega_1}/[-(\omega_0 - \omega_1)] + P_{\omega_1}/\omega_1 = 0$$

or

$$P_{\omega_0-\omega_1}/(\omega_0 - \omega_1) = P_{\omega_2}/\omega_2 = P_{\omega_1}/\omega_1$$

In a similar way, the second equation yields

$$P_{\omega_2}/\omega_2 = -P_{\omega_0}/\omega_0$$

Thus, we obtain the following proportions

$$P_{\omega_1}/\omega_1 = P_{\omega_2}/\omega_2 = -P_{\omega_0}/\omega_0$$

coinciding with expression (10.55) if ω_0 is replaced by ω_p.

From this analysis it is clear that a nonlinear capacitor can be used for a spectrum conversion accompanied by the pumping of energy from one source into another. Thus, if ω_1 is the frequency of the received signal and ω_0 is the frequency of the pumping oscillator, one can separate the combination frequency $\omega_2 = \omega_0 - \omega_1$ with simultaneous power amplification at this frequency. It should be recalled that if a resistive nonlinear element is used, the conversion of the signal frequency (see Sec. 8.10) is not accompanied by the pumping of energy from the local oscillator.

10.6. PRINCIPLE OF PARAMETRIC AMPLIFICATION OF OSCILLATIONS

In Sec. 10.1 it was shown that with respect to the signal that is small in comparison with the control oscillation, the nonlinear capacitance together with the pumping oscillator can be replaced by a linear time-varying capacitance. Putting aside the method for modulating capacitance (or inductance), one can speak about energy exchange between the signal and the energy-storing parametric element.

A graphic example of the energy exchange that occurs when changing a capacitance is the well-known model of a charged capacitor

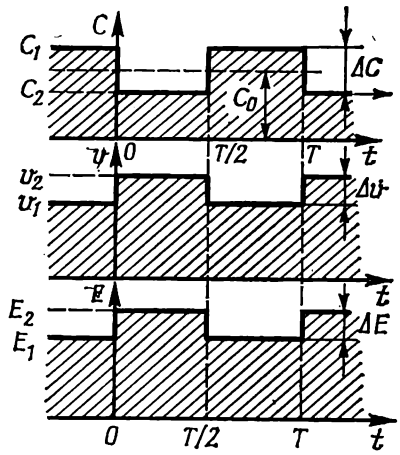


Fig. 10.11. Voltage across a capacitor and the stored energy in the case of stepwise changes in the capacitance. The capacitor charge is constant

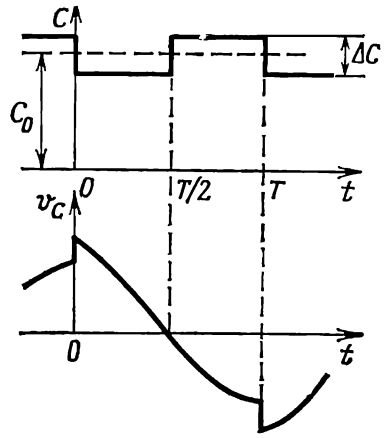


Fig. 10.12. Increase in the voltage amplitude across a capacitor at the moments the capacitance decreases stepwise

whose plates can be mechanically moved wider apart or closer together. The ponderomotive force of the electric field of the capacitor tends to move the capacitor plates closer together (regardless of the voltage polarity). Consequently, to move them wider apart, i.e., to reduce the capacitance, it is necessary to do work that increases the energy stored in the capacitor. On the contrary, when moving the plates closer together, some energy of the capacitor field is converted into mechanical energy. It is not difficult to establish the relation between the relative change in the capacitance of the capacitor and the change in the energy stored in this capacitor.

Let us consider a capacitor whose capacitance $C(t)$ varies in a stepwise manner, as shown in Fig. 10.11. Let the capacitor get a charge q which then remains constant (the open-circuited capacitor is supposed to be loss-free). At the moments the capacitance instantaneously decreases by an amount $\Delta C = C_1 - C_2$, the voltage

across the capacitor increases by an amount $\Delta v = v_2 - v_1 = q\,(1/C_2 - 1/C_1) = q\,(C_1 - C_2)/C_1C_2 = v_1\,\Delta C/C_2$, while the energy increases by an amount $\Delta E = E_2 - E_1 = (q^2/2)\,(1/C_2 - 1/C_1) = (q^2/2C_1)\,\Delta C/C_2 = E_1\,\Delta C/C_2$. At the moments the capacitance increases stepwise, the voltage and energy decrease by Δv and ΔE, respectively.

In the first case, the additional energy ΔE is taken from the device that effects the stepwise decrease of the capacitance, while in the second, the inverse conversion of energy takes place. On the average, during the time T, the capacitor energy remains constant.

It is not the case if the capacitor charge is a function of time such that the capacitance decreases at the moments $q(t)$ passes through its maxima and increases at the moments $q(t)$ passes through the minima (better still, at the moments it passes through zero). Such operating conditions can be realized by inserting the capacitor $C(t)$ into a tuned circuit whose oscillation frequency is half the variation frequency of the capacitance. Assume that a high-Q tank circuit is excited by a signal $e(t) = E \cos \omega t$ whose frequency is the same as the resonance frequency $\omega_r = 1/\sqrt{LC_0}$ of the tank circuit, ω_r satisfying the condition $\omega_r = 2\pi f_r = 2\pi/2T$, where T is the complete variation cycle of $C(t)$. In this case, the voltage across the tuned circuit can be represented in the form of an oscillation very close to harmonic but incremented at the moments the capacitance $C(t)$ drops (Fig. 10.12). This is equivalent to increasing the signal power. If the energy increment due to a single capacitance decrement ΔC does not exceed the expenditure of energy during the time T, the parametric circuit is stable, otherwise the parametric excitation of oscillations occurs. Thus, by controlling the quantity $\Delta C/C_0$, i.e., the depth of modulation of the parameter C, one can effect both parametric amplification of a signal and parametric excitation of oscillations.

The stepwise variation of $C(t)$ is technically difficult to realize and is not used in practice. It is much more simple to modulate it according to a harmonic law, only one has to observe the main principle: to decrease the capacitance of a capacitor when its charge (voltage) is a maximum and to increase it when the charge is a minimum.

In the next sections we shall consider equivalent circuits of capacitive and inductive energy-storing elements whose parameters C and L vary in time according to a harmonic law and establish the basic relations which are necessary for the analytic description of the processes of amplification and excitation of oscillations in parametric circuits.

10.7. EQUIVALENT CIRCUIT OF A HARMONICALLY VARYING CAPACITANCE OR INDUCTANCE

Let a signal voltage $e(t) = E \cos \omega t$ be applied to an electronically-controlled capacitance (a varicap) and let it be required to find the current flowing through the signal source on the premise that the control oscillation (pumping oscillator) frequency ω_0 is twice the signal frequency ω. The method of obtaining a periodically-varying capacitance is illustrated in Fig. 10.13*a*. A control

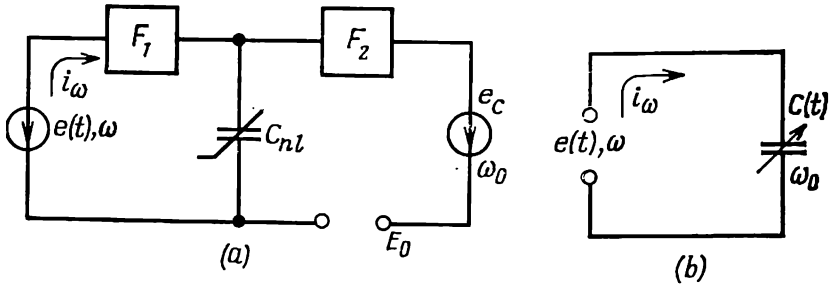


Fig. 10.13. (*a*) Effect of the pumping and signal voltages on a linear capacitor and (*b*) equivalent circuit for a low-level signal

(pumping) voltage $e_c = E_c \cos(\omega_0 t + \gamma)$ superimposed on a d-c voltage E_0 is applied to a nonlinear capacitance C_{nl}.

A filter F_1 does not allow the current of frequency ω_0 to flow to the signal-source circuit, while a filter F_2 does not allow the current of signal frequency ω (and those of frequencies close to ω) to flow to the pumping circuit.

Let us impose the condition $E \ll E_c$. In this case, as stated in Sec. 10.1, we may neglect the change in the capacitance due to the action of the signal and consider that the variation of the capacitance is defined solely by the control voltage. Using formula (10.5), we set (for $m \ll 1$)

$$C(t) = C_0/[1 + m \cos(\omega_0 t + \gamma)] \approx C_0 - \Delta C \cos(\omega_0 t + \gamma) \quad (10.57)$$

where

$$\Delta C = mC_0 = \frac{E_c}{2(e_{con} + E_0)} C_0 \quad (10.58)$$

and γ is the epoch angle.

Shown in Fig. 10.13*b* is an equivalent *linear* parametric circuit (the pumping circuit is not shown).

Let us determine the total current through the capacitance $C(t)$ using general expression (10.7):

$$i(t) = [C_0 - \Delta C \cos(\omega_0 t + \gamma)][-\omega E \sin \omega t] + E \cos \omega t \; \omega_0 \Delta C$$

$$\times \sin(\omega_0 t + \gamma) = -\omega C_0 E \sin \omega t + \frac{1}{2}(\omega_0 + \omega) \Delta C E \sin[(\omega_0 + \omega) t + \gamma]$$

$$+ \frac{1}{2}(\omega_0 - \omega) \Delta C E \sin[(\omega_0 - \omega) t + \gamma] \quad (10.59)$$

Note that this expression can be obtained directly from (10.51) on setting $b_1 = C_0$, $E_1 = E$, $\theta_1 = \pi/2$, $E_2 = E_c$, $\theta_2 = \pi/2 + \gamma$, $b_2E_2 = b_2E_c = \Delta C/2$ and omitting the term with the coefficient E_1^2 (for it is very small) and also the terms independent of E_1.

The frequency $\omega_0 + \omega \approx 3\omega$ lies outside the passband of the filter F_1, therefore, the current in the signal-source circuit is a sum of two currents: one of frequency ω and the other of combination frequency $\omega_0 - \omega$ that is close to ω (since $\omega_0 \approx 2\omega$). The first of these currents is shifted in phase by 90° with respect to $e(t) = E \cos \omega t$, so it cannot be taken account of by any conductance,

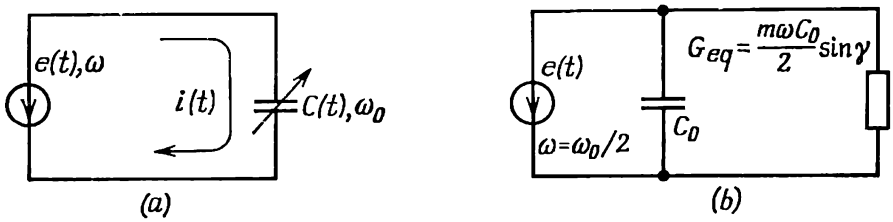


Fig. 10.14. (*a*) Parametric capacitive network and (*b*) equivalent circuit for a signal whose frequency is half the pumping frequency

whether positive or negative. From the standpoint of obtaining amplification effect, the combination oscillation of the difference frequency $\omega_0 - \omega$ is of special interest, particularly in the case $\omega_0 = 2\omega$. In this case, the current of frequency ω is

$$i_{\omega_0-\omega}(t) = i_\omega(t) = \frac{1}{2}(\omega_0 - \omega)\,\Delta CE \sin[(\omega_0 - \omega)t + \gamma]$$

$$= \frac{1}{2}\omega\Delta CE \sin(\omega t + \gamma) = \frac{1}{2}\omega\Delta CE \cos[\omega t - (\pi/2 - \gamma)]$$

$$= I_\omega \cos[\omega t - (\pi/2 - \gamma)] \qquad (10.60)$$

For the signal-source e.m.f. $e(t) = E \cos \omega t$ and the current $i_\omega(t)$ defined by expression (10.60), the power output of the signal source is

$$P_\omega = \frac{EI_\omega}{2}\cos\left(\frac{\pi}{2} - \gamma\right) = \frac{1}{2}\omega\Delta C\frac{E^2}{2}\sin\gamma = G_{eq}\frac{E^2}{2}$$

where the symbol

$$G_{eq} = \frac{\omega\Delta C}{2}\sin\gamma = \frac{m\omega C_0}{2}\sin\gamma \qquad (10.61)$$

stands for the equivalent conductance accounting for the consumption of the signal-source power.

Thus, we come to the equivalent circuit shown in Fig. 10.14*b*, which corresponds to the parametric circuit of Fig. 10.14*a*. The combination frequency $\omega_0 + \omega = 3\omega$ is not taken account of in

this circuit, while the frequency $\omega_0 - \omega$ coincides with ω. As a result, with respect to the signal source, the parametric circuit of Fig. 10.14*a* is reduced to a circuit with *constant parameters*. The periodic variation of $C(t)$ with a frequency $\omega_0 = 2\omega$ results only in the development of the conductance G_{eq} shunting the constant capacitance C_0.

Let us consider the following three typical operating conditions: $\gamma = 0$, $\gamma = \pi/2$, and $\gamma = -\pi/2$ (Fig. 10.15). In the first case ($\gamma = 0$),

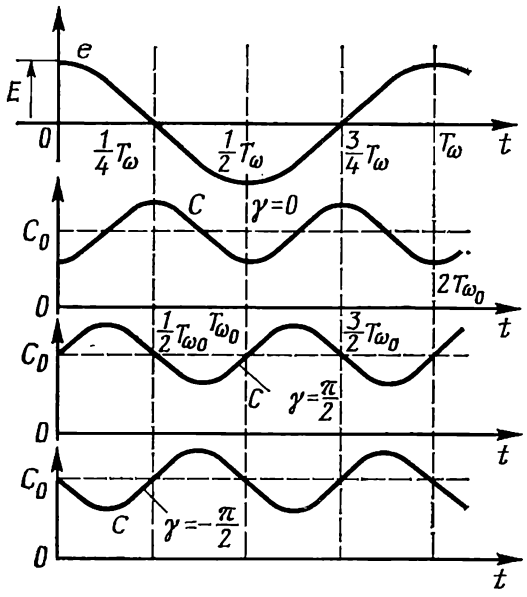


Fig. 10.15. Voltage across a capacitor and the laws governing its variation for various epoch angles

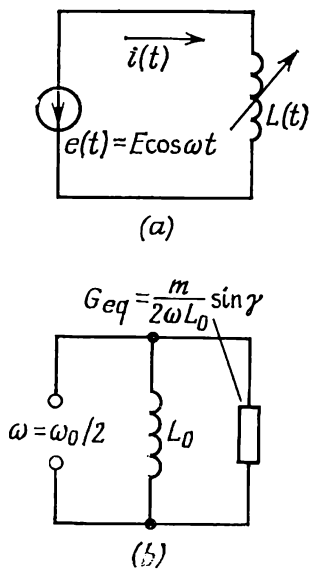


Fig. 10.16. (*a*) Parametric inductive network and (*b*) equivalent circuit for a signal whose frequency is half the pumping frequency

$C(t)$ is modulated so that during the period $T_{\omega_0} = 2\pi/\omega_0$ (as well as during the period $T_\omega = 2\pi/\omega$) the change in the amount of energy stored in the capacitance is zero. In this case, $G_{eq} = 0$.

In the second case ($\gamma = \pi/2$), the maximum rate of rise of $C(t)$ occurs at the moments the voltage $e(t)$ passes through its maxima; in this case, some of the energy stored in the capacitance goes to the device that varies the capacitance. With respect to the signal source, this is equivalent to shunting the constant capacitance C_0 by a positive conductance $G_{eq} = (m/2)\omega C_0$.

Finally, in the third case ($\gamma = -\pi/2$) where $C(t)$ decreases in the region $e(t) = E$ and increases in the region $e(t) = 0$, the conductance $G_{eq} = -(m/2)\omega C_0$ is negative and equal to $-(m/2)\omega C_0$.

This result agrees with the results of the qualitative analysis of parametric amplification (see the end of the preceding section). The negative conductance G_{eq} takes account of the energy supply from the pumping oscillator to the circuit containing $C(t)$. In the given example of an electronically-controlled capacitance, the increase of the amount of energy stored in the capacitance is due to the work done during the decrease in the capacitance by the pumping oscillator in overcoming electric field forces to cause electrons and holes to move through the potential barrier in the region of the barrier layer.

Similar results can easily be obtained for a periodically-varying inductance $L(t)$. Proceeding from the circuit shown in Fig. 10.16*a*, where the inductance is being varied according to

$$L(t) = L_0 [1 + m \cos(\omega_0 t + \gamma)] \tag{10.62}$$

we find the current by means of relation (10.11) (for $m \ll 1$):

$$i(t) = \frac{1}{L(t)} \int e(t)\, dt = \frac{1}{\omega L_0 [1 + m \cos(\omega_0 t + \gamma)]} E \sin \omega t$$

$$\approx \frac{E}{\omega L_0} [1 - m \cos(\omega_0 t + \gamma)] \sin \omega t$$

$$= E \left\{ \frac{1}{\omega L_0} \sin \omega t - \frac{m}{2\omega L_0} \sin [(\omega + \omega_0) t + \gamma] \right.$$

$$\left. - \frac{m}{2\omega L_0} \sin [(\omega - \omega_0) t - \gamma] \right\}$$

With $\omega_0 = 2\omega$, the current of frequency ω is

$$i_\omega(t) = E \frac{1}{\omega L_0} \sin \omega t + E \frac{m}{2\omega L_0} \sin(\omega t + \gamma)$$

$$= E \frac{1}{\omega L_0} \sin \omega t + E \frac{m}{2\omega L_0} \cos \left[\omega t - \left(\frac{\pi}{2} - \gamma \right) \right]$$

The first term has no effect on the power consumption, while the second, which is shifted by an angle $\pi/2 - \gamma$ relative to the signal e.m.f., defines the power consumption to be

$$P = \frac{m}{2\omega L_0} \frac{E^2}{2} \cos \left(\frac{\pi}{2} - \gamma \right) = \frac{m}{2\omega L_0} \frac{E^2}{2} \sin \gamma = G_{eq} \frac{E^2}{2}$$

where

$$G_{eq} = \frac{m}{2\omega L_0} \sin \gamma$$

is the equivalent conductance.

Thus, with $\omega_0 = 2\omega$, one gets the equivalent circuit shown in Fig. 10.16*b*. The phase relations between $e(t) = E \cos \omega t$, $i(t) = (E/\omega L_0) \sin \omega t$, and the inductance $L(t)$ that varies according

to (10.62) are illustrated in Fig. 10.17 plotted for $\gamma = -\pi/2$. In this case, the conductance G_{eq} will be negative ($-m/2\omega L_0$) if the function $L(t)$ decreases when the current $i(t)$ passes through its maxima and increases when it passes through zero.

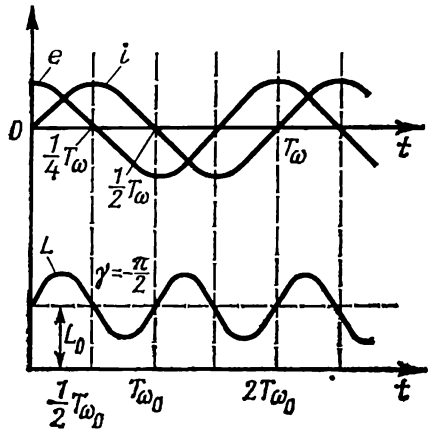


Fig. 10.17. Voltage across and current in a coil whose inductance decreases when the current passes through its maxima

The energy is supplied to the circuit at the expense of the work done during the decrease in the inductance by the pumping device in overcoming magnetic field forces tending to move the turns of the inductance coil closer together and thus increase the inductance of the coil.

10.8. SINGLE-TUNED PARAMETRIC AMPLIFIER

From the previous section it follows that by inserting a variable capacitance or inductance into a tuned circuit one can effect the amplification of oscillations, provided the parameter in question is made to vary appropriately. Figure 10.18*a* shows a simple single-tuned parametric amplifier with a variable capacitance. The non-

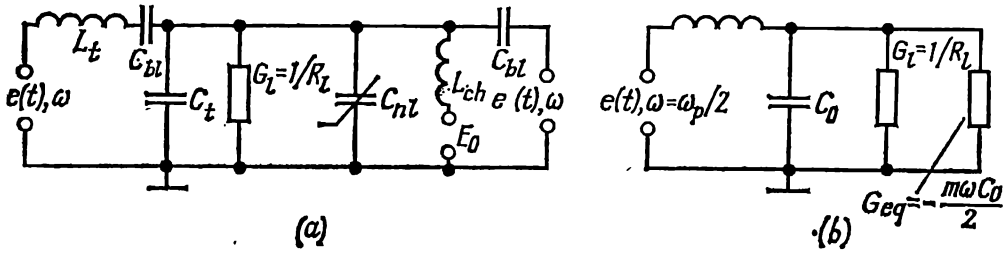


Fig. 10.18. (*a*) Single-tuned parametric amplifier and (*b*) its equivalent circuit

linear capacitor C_{nl} is acted upon by two voltages — a signal voltage of frequency ω and a pumping (control) voltage of frequency ω_p.

The blocking capacitors C_{bl} protect the pumping oscillator and the signal source from the d-c voltage E_0 used for setting the operating point on the volt-farade characteristic of the varicap. The choke L_{ch} prevents the flow of the r-f currents of frequencies ω and ω_p to the source of E_0.

First, let us consider the operation of the amplifier in the case where the condition

$$\omega = \omega_p/2$$

is strictly satisfied.

In this, so-called "synchronous" operation, the combination frequency $\omega_p - \omega$ coincides with the frequency ω, so that in the tuned circuit there exists only the current of frequency ω. The equivalent circuit for synchronous operation in the case $\gamma = -\pi/2$ corresponding to a negative conductance G_{eq} is shown in Fig. 10.18*b*.

The symbol C_0 stands for the sum of the tank capacitance C_t and the average capacitance (corresponding to the d-c voltage E_0) of the varicap.

To simplify the analysis, the source of e.m.f. $e(t)$, connected in series into the tuned circuit, is replaced in Fig. 10.19 by a current

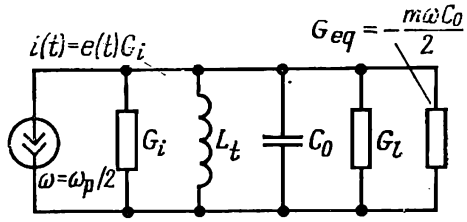


Fig. 10.19. Single-circuit parametric amplifier (to the circuit of Fig. 10.18*a*)

generator which is connected in parallel with the tuned circuit and shunted by the internal conductance G_i. The load conductance G_l also includes the conductance that accounts for the power losses in the tank elements. Shunting the load conductance G_l by the negative conductance $G_{eq} = (\omega \Delta C/2) \sin\gamma = -\omega \Delta C/2 = -m\omega C_0/2$ reduces the total conductance, thereby increasing the Q-factor of the tuned circuit. Thus, we have an amplification effect.

Now, let us set up an expression for the amplification factor in the form of the ratio between the signal power at the amplifier output and the maximum power obtainable without parametric modulation.

As is known, the power dissipated in the load conductance (in the absence of amplification) reaches its maximum when $G_l = G_i$.

In this case, the signal power is

$$P_s = \frac{1}{2} \frac{I^2}{4G_l} = \frac{1}{2} \frac{I^2}{4G_i}$$

(I is the generator current amplitude).

When the additional conductance G_{eq} is connected, the output voltage will be $E = I/(G_i + G_l + G_{eq}) = I/(2G_l + G_{eq})$ and the power dissipated in the load conductance will be given by

$$P'_s = \frac{1}{2} E\,(EG_l) = \frac{1}{2} G_l E^2 = \frac{1}{2} G_l \frac{I^2}{(2G_l + G_{eq})^2} = \frac{1}{2} \frac{I^2}{4G_l} \frac{1}{(1 + G_{eq}/2G_l)^2}$$

Whence the power amplification factor

$$K_P = P'_s/P_s = 1/(1 + G_{eq}/2G_l)^2 \qquad (10.63)$$

It should be recalled that G_{eq} has a *negative value.*

From this expression directly follows the condition for the stability of a parametric amplifier (in synchronous operation):

$$|G_{eq}| < 2G_l \quad \text{or} \quad m\omega C_0/2 < 2G_l \tag{10.64}$$

whence the critical value of the parametric modulation factor is

$$m_{cr} = 2\,(2G_l/\omega C_0) = 2/Q_{eq} \tag{10.65}$$

where Q_{eq} is the Q-factor of the tuned circuit with due regard for G_i and $G_l = G_i$.

It should be noted that when $G_{eq} = -G_l$, i.e., when parametric modulation compensates for the losses in G_l only, the power amplification factor is no more than four.

In practice, when amplifying a real signal whose phase is unknown and whose frequency may vary within some range, synchronous operation cannot be ensured.

Suppose that the signal frequency ω is not $\omega_p/2$ exactly, but $\omega_p/2 + \Omega$, where Ω is a small deviation not extending beyond the passband of the tuned circuit. Then, the combination frequency will be

$$\omega_p - \omega = \omega_p - (\omega_p/2 + \Omega) = \omega_p/2 - \Omega$$

In this case, within the passband of the tuned circuit there will be two oscillations — one of frequency $\omega_p/2 + \Omega$ (useful signal) and the other of frequency $\omega_p/2 - \Omega$ (combination frequency).

The relation between the amplitudes of these two oscillations depends on the depth m of capacitance modulation and the value of Ω. A detailed analysis [2] shows that for m close to critical [see formula (10.65)] and a comparatively small detuning Ω, the amplitudes of both oscillations are approximately the same. There develop beats causing amplitude pulsations and changes in the phase of the resultant oscillation. But even if there is a difference between the frequencies ω and $\omega_p/2$, the average oscillation power per beat period can be shown to be higher than in the absence of parametric action, i.e., even in this, so-called *biharmonic operation*, the signal is amplified. However, such an operation of the amplifier is not always acceptable.

The circuit discussed in the next section is free from the disadvantages inherent in the single-tuned parametric amplifier.

10.9. DOUBLE-FREQUENCY PARAMETRIC AMPLIFIER

Figure 10.20 shows the circuit diagram of a double-frequency or double-tuned parametric amplifier. The first, the signal tank circuit, is tuned to the central frequency of the signal spectrum (resonance frequency $\omega_{r1} \approx \omega_1$), while the second, the "idle" tank circuit, is tuned to the frequency ω_{r2} differing substantially from ω_{r1}.

The pumping frequency is selected proceeding from the condition

$$\omega_p = \omega_{r1} + \omega_{r2} \tag{10.66}$$

When selecting the frequency ω_{r2}, the basic requirement is that the signal frequency ω_1 should be outside the passband of the auxiliary tank circuit. On the other hand, the combination frequency $\omega_2 = \omega_p - \omega_1$ must lie outside the operating band of the signal circuit.

If these conditions are satisfied, across the signal tank circuit there will exist only the voltage of frequency ω_1 and across the

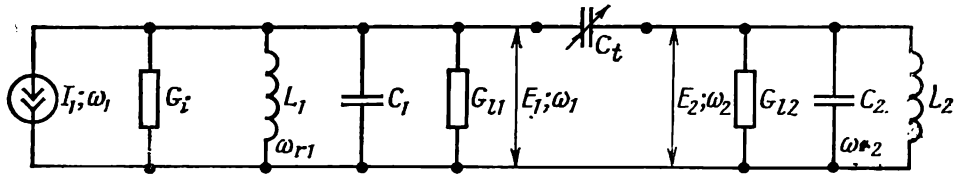


Fig. 10.20. Double-frequency parametric amplifier

auxiliary tank circuit, the voltage of frequency ω_2. Assuming that the amplitudes E_1 and E_2 of these voltages are low compared to E_p, we may replace the nonlinear capacitance C_{nl} together with the pumping oscillator by a *linear* parametric capacitance $C(t)$ varying with frequency ω_p, as it was done in Sec. 10.7. In this case, under the action of the signal voltage $e_1(t) = E_1 \cos(\omega_1 t + \theta_1)$, in the circuit of the variable capacitance $C(t) = C_0 - \Delta C \cos(\omega_p t + \omega_p)$ there arises (in addition to other components which are of no interest here) a current

$$i_{\omega_p - \omega_1}(t) = i_{\omega_2}(t) = \frac{1}{2}(\omega_p - \omega_1)\Delta C E_1 \sin[\omega_p t + \theta_p - (\omega_1 t + \theta_1)]$$

$$= \frac{1}{2}\omega_2 \Delta C E_1 \sin[\omega_2 t + (\theta_p - \theta_1)] = I_{\omega_2} \sin[\omega_2 t + (\theta_p - \theta_1)] \tag{10.67}$$

[see expression (10.59)]. Here, $I_{\omega_2} = \frac{1}{2}\omega_2 \Delta C E_1$.

The current $i_{\omega_2}(t)$ produces a voltage drop across the impedance $Z_2(i\omega_2) = Z_2(\omega_2)\, e^{i\varphi_Z}$ of the idle tank circuit, which is given by

$$I_{\omega_2} Z_2(\omega_2) \sin(\omega_2 t + \theta_p - \theta_1 + \varphi_Z)$$

$$= \frac{1}{2}\omega_2 \Delta C Z_2(\omega_2) E_1 \sin(\omega_2 t + \theta_p - \theta_1 + \varphi_Z)$$

As in Sec. 10.5, let us write the equivalent e.m.f. acting on the capacitance $C(t)$ in the form

$$e_2(t) = E_2 \cos(\omega_2 t + \theta_2) = -\frac{1}{2}\omega_2 \Delta C Z_2(\omega_2) E_1 \sin(\omega_2 t + \theta_p - \theta_1 + \varphi_Z)$$

$$= E_2 \cos(\omega_2 t + \theta_p - \theta_1 + \varphi_Z + \pi/2)$$

where

$$E_2 = \frac{1}{2}\,\omega_2\,\Delta C Z_2(\omega_2)\,E_1$$

The combination current $i_{\omega_p-\omega_2}(t)$ due to this e.m.f. will, by analogy with (10.67), be

$$\begin{aligned} i_{\omega_p-\omega_2}(t) = i_{\omega_1}(t) &= \frac{1}{2}(\omega_p-\omega_2)\,\Delta C E_2 \\ &\times \sin[(\omega_p-\omega_2)\,t+\theta_p-(\theta_p-\theta_1+\varphi_Z+\pi/2)] \\ &= \frac{1}{2}\omega_1\Delta C E_2 \sin(\omega_1 t+\theta_1-\varphi_Z-\pi/2) \\ &= -\frac{1}{2}\omega_1\Delta C E_2\cos(\omega_1 t+\theta_1-\varphi_Z) \end{aligned} \qquad (10.68)$$

Taking into account the above expression for E_2, (10.68) can be written in the form

$$i_{\omega_1}(t) = -(\Delta C/2)^2\,\omega_1\omega_2 Z_2(\omega_2)\,E_1\cos(\omega_1 t+\theta_1-\varphi_Z)$$

As is seen, with respect to the signal tank circuit, the nonlinear capacitance C_{nl} together with the pumping oscillator and the idle tank circuit can be replaced by the equivalent conductance accounting for the current $i_{\omega_1}(t)$ thus found.

The complex amplitude of this current

$$\mathbf{I}_{\omega_1} = -(\Delta C/2)^2\,\omega_1\omega_2 Z_2(\omega_2)\,e^{-\varphi_Z}E_1 e^{i\theta_1}$$

On the other hand, the complex amplitude of the voltage $e_1(t) = E_1\cos(\omega_1 t+\theta_1)$ across the signal tank circuit is $\mathbf{E}_1 = E_1 e^{i\theta_1}$.

Consequently, the conductance shunting the signal tank circuit will be

$$\begin{aligned} G_{eq}(i\omega_1) = \frac{\mathbf{I}_{\omega_1}}{\mathbf{E}_1} &= -\left(\frac{\Delta C}{2}\right)^2\omega_1\omega_2 Z_2(\omega_2)\,e^{-i\varphi_Z} \\ &= -\left(\frac{\Delta C}{2}\right)^2\omega_1\omega_2 Z_2^*(i\omega_2) = -\left(\frac{mC_0}{2}\right)^2\omega_1\omega_2 Z_2^*(i\omega_2) \end{aligned} \qquad (10.69)$$

where $Z_2^*(i\omega_2) = Z_2(\omega_2)\,e^{-i\varphi_Z}$ is the complex conjugate of the function $Z_2(i\omega_2)$.

At resonance, when $\omega_1 = \omega_{r1}$ and consequently, $\omega_2 = \omega_{r2}$, the resistance of the auxiliary tank circuit is $R_{l2} = 1/G_{l2}$ and formula (10.69) assumes the form

$$G_{eq}(\omega_{r1}) = -(mC_0/2)^2\omega_1\omega_2 R_{l2} \qquad (10.69')$$

In the equivalent circuit shown in Fig. 10.21, the elements located to the left of the dashed line correspond to the signal tank circuit of the amplifier, while those to the right correspond to the nonlinear capacitance together with the auxiliary tank circuit. This circuit is in fact the same as that of the single-tuned amplifier (see Fig. 10.19).

The only difference is in the method of determining the equivalent negative conductance.

It should be noted that the above relations could be obtained in a simpler way, using expressions (10.44) through (10.56). The details associated with the determination of the combination oscillations $i_{\omega_p-\omega_1}(t)$ and $i_{\omega_p-\omega_2}(t)$ are given with a view to attracting atten-

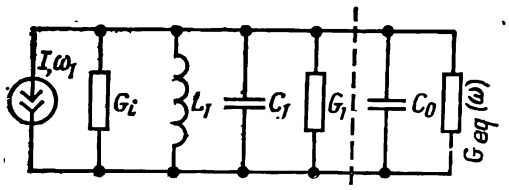


Fig. 10.21. Equivalent circuit of a double-tuned parametric amplifier

tion to the following advantages of the double-tuned parametric amplifier:

(a) the equivalent negative conductance and consequently, power amplification, are *independent of the phase of the pumping voltage*;

(b) there is no need in keeping to a certain definite relation between the frequencies ω_1 and ω_p.

Both these properties of the double-tuned parametric amplifier are explained by the fact that in expression (10.68) the phase of the combination current $i_{\omega_p-\omega_2}$, that determines the character of the equivalent conductance G_{eq}, is in fact the difference between the phases of the pumping voltage $e_p(t)$ and $e_2(t)$. The first of these phases is $(\omega_p t + \theta_p)$ and the second, $(\omega_2 t + \theta_p - \theta_1)$ (without taking into account φ_Z and $\pi/2$). When the difference is formed, θ_p is rejected, and the difference frequency $\omega_p - \omega_2$ is always the same as the signal frequency (because $\omega_2 = \omega_p - \omega_1$).

The amplification factor of the double-tuned parametric amplifier at the resonance frequency ($\omega_1 = \omega_{r1}$) can be defined from an expression similar to formula (10.63):

$$K_P = 1/(1 + G_{eq}/2G_{l1})^2 \qquad (10.70)$$

where G_{eq} is calculated by formula (10.69) and G_{l1} is the load conductance of the signal tank circuit.

When the signal frequency ω_1 deviates from the resonance frequency ω_{r1} and, respectively, the frequency ω_2 deviates from ω_{r2}, the magnitude of the impedance $Z(i\omega_2)$ decreases, this leading to a decrease in the magnitude of G_{eq} and consequently, in the power amplification factor.

Using expression (10.69), we can calculate the amplitude-frequency characteristic and the passband of the double-frequency parametric amplifier.

In this case, the condition for the stability of the amplifier may be written in the form

$$|G_{eq}| = (mC_0/2)^2 \omega_1 \omega_2 R_{l2} < 2G_{l1}$$

or

$$m < 2\sqrt{2}\sqrt{G_{l1}/R_{l2}\omega_1\omega_2 C_0^2} \qquad (10.71)$$

Let us consider the energy balance in the double-frequency parametric amplifier, depending on the relation between the frequencies ω_1 and ω_2.

Let the frequency ω_1 and the power P_s of the signal at the amplifier input be specified. Since the magnitude of the negative quantity

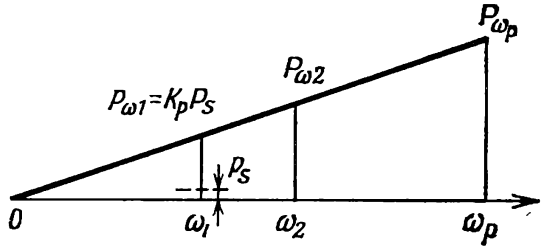


Fig. 10.22. Power relations at different frequencies in a double-tuned parametric amplifier

G_{eq} increases with an increase in the auxiliary frequency ω_2 [see formula (10.69)], the power amplification factor K_P also increases [see formula (10.70)]. The signal power at the amplifier output will be $P\omega_1 = K_P P_s$.

To find the required power P_{ω_p} of the pumping oscillator and the power P_{ω_2} dissipated in the auxiliary tank circuit, let us use the Manly-Raw theorem. Using expressions (10.55), we can write the following relations:

$$P_{\omega_2} = \frac{\omega_2}{\omega_1} P_{\omega_1}, \quad P_{\omega_p} = \frac{\omega_p}{\omega_1} P_{\omega_1} = P_{\omega_1} + P_{\omega_2}$$

(The minus sign in the last expression is omitted, since it is obvious that this power is taken from the pumping oscillator.) The relations between the powers P_s, P_{ω_1}, P_{ω_2}, and P_{ω_p} are illustrated in Fig. 10.22. From this plot it is clear that for $\omega_2 > \omega_1$, the power dissipated in the auxiliary tank circuit is higher than in the signal one. Thus, although the power P_{ω_1} increases with an increase in the frequency ω_2, the distribution of the power taken from the pumping oscillator is changed in favour of the frequency ω_2. In spite of this, the condition $\omega_2 > \omega_1$ is often employed, because, when amplifying weak signals, the ratio of P_{ω_1} to P_s, i.e., the gain

K_P, is of greater importance than the degree of utilization of the power P_{ω_p}.

To illustrate the quantitative relations in the double-frequency parametric amplifier, let us consider the following example.

Let it be required to amplify a signal of frequency $f_1 = 30$ MHz and spectrum width $2\Delta f_0 = 100$ kHz.

The initial data of the first (signal) tank circuit: characteristic impedance $\rho_1 = 100\ \Omega$. internal resistance of the signal source shunting the circuit, $R_i = 5$ kΩ, load resistance $R_{l1} = 5$ kΩ.

The initial data of the second (auxiliary) tank circuit: resonance frequency $f_{r2} = 60$ MHz, characteristic impedance $\rho_2 = 50\ \Omega$, load resistance $R_{l2} = 5$ kΩ.

Before calculating the required variation of the varicap capacitance, let us find the maximum conductance G_{eq} which can be connected to the signal tank circuit to suit the specified signal-spectrum width of $2\Delta f_0$.

The Q-factor of the signal tank circuit (if shunted by a negative conductance) apparently should not exceed

$$Q_1 \leqslant f_1/2\Delta f_0 = 30 \times 10^6/(100 \times 10^3) = 300$$

At $\rho_1 = 100\ \Omega$, the resultant conductance shunting the first tank circuit must not be less than

$$G_i + G_{l1} + G_{eq} \geqslant 1/\rho_1 Q_1$$

whence

$$G_{eq} \geqslant 1/\rho_1 Q_1 - (G_i + G_{l1}) = 1/\rho_1 Q_1 - 2G_{l1} = -367 \times 10^{-6}\ \text{S}$$

Substituting G_{eq}, ω_1, ω_2, and R_{l2} into formula (10.69′), we find

$$mC_0/2 = \Delta C/2 = \sqrt{|G_{eq}|/\omega_1\omega_2 R_{l2}} \approx 3 \times 10^{-12}\ \text{F}$$

whence

$$\Delta C = 6 \times 10^{-12}\ \text{F} = 6\ \text{pF}$$

The required value of ΔC can be provided by an ordinary varicap. At present, there are varicaps which can provide a change in capacitance of up to 30 pF.

The amplification factor can be calculated by formula (10.70):

$$K_P = 1\Big/\left(1 - \frac{367 \times 10^{-6}}{2 \times 200 \times 10^{-6}}\right)^2 \approx 147$$

In conclusion, let us note the main advantages and disadvantages of the parametric amplifier.

An important advantage of the parametric amplifier is its low noise level as compared with the transistor or the vacuum-tube amplifier. In Sec. 7.2 it was noted that the main source of noise in transistor and vacuum-tube amplifiers is the shot effect due to the random transfer of discrete charges of electrons and holes (in a transistor). In the parametric amplifier, a similar effect occurs in the

device modulating the parameter. For example, in the varicap, the capacitance variation depends on the movement of electrons and holes. However, the flow intensity of the charge carriers in a varicap is much lower than in a transistor or an electron tube. In the latter two, the flow intensity defines the power of the useful signal in the load circuit, while in the varicap, this intensity defines only the effect of modulation of the parameter. The influence of the shot effect is reduced here to such an extent that the noise level in the parametric amplifier is determined mainly by the thermal noise. In this connection, parametric diodes are often cooled down to (5 to 10) K.

The complicacy of decoupling of the pumping and signal tank circuits is a disadvantage of the parametric amplifier.

In the system shown in Fig. 10.18*a*, which is typical of parametric amplifiers of a metric band, the decoupling is effected by means of blocking capacitors and chokes. In the SHF band, where parametric amplifiers are widely used, fairly complex constructions have to be employed, combining in a single unit a double-frequency oscillatory circuit in the form of hollow resonators, a varicap, and special decoupling elements (circulator, directional coupler, absorber, rejection filter). These questions are considered in special courses.

10.10. FREQUENCY CONVERSION BY MEANS OF A NONLINEAR REACTIVE ELEMENT

The above-discussed double-frequency parametric amplifier can be used as a frequency converter if an oscillation of a combination frequency is isolated at the amplifier output. For the circuit shown in Fig. 10.20, this frequency is equal to $\omega_2 = \omega_p - \omega_1$. For $\omega_2 > > \omega_1$ the power that can be drawn at the frequency ω_2 exceeds that at the frequency ω_1 of the input signal by a factor of ω_2/ω_1 (Fig. 10.22). Under such conditions, the signal spectrum is reversed, as in the case discussed in Sec. 8.10, where $\omega_{lo} > \omega_s$ (the pumping oscillator in the circuit of Fig. 10.20 plays the part of the local oscillator).

When operating in the SHF band, the realization of pumping oscillators operating at frequencies much higher than the signal frequency is a difficult problem. In addition, further processing of the signal in the case of the "upward" frequency conversion is complicated. Therefore, widely used in practice are operating conditions where the frequencies ω_1 and ω_2 are close. In this case, amplification is obtained by using the regeneration effect and not by increasing the ratio ω_2/ω_1.

In the long, short, and ultrashort wave ranges, where it is possible to use high-power pumping oscillators operating at frequencies much higher than the signal frequency, parametric amplification can be

combined with the "upward" frequency conversion. The spectrogram of such an amplifier-converter is shown in Fig. 10.23. The amplified oscillation is separated at the combination frequency $\omega_2 = \omega_p + \omega_1$. The power of this oscillation is equal to the sum of the input-signal power P_s and the power P_p drawn from the pumping oscillator. It should be borne in mind that in the given case we have no regeneration effect and there is no amplification at the frequency ω_1.

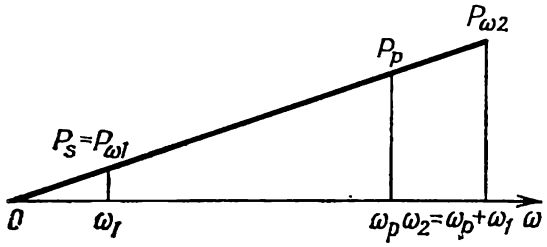


Fig. 10.23. Power relations at different frequencies in a parametric amplifier-frequency converter

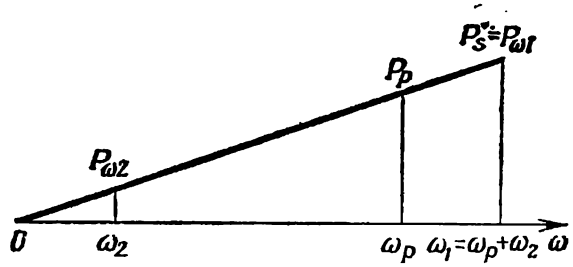


Fig. 10.24. Power relations in a parametric nonregenerative frequency converter

A nonregenerative frequency conversion can also be effected in a circuit with a nonlinear capacitance by using a pumping frequency $\omega_p < \omega_1$ and separating the difference frequency $\omega_2 = \omega_1 - \omega_p$ in the load (Fig. 10.24). In this case, there is no amplification at all. The input signal is the only source of energy. This energy is consumed both in the circuit containing the load resistor (at the frequency ω_2) and in that containing the pumping oscillator. When using the Manly-Raw theorem (see Sec. 10.5), this result stems from the fact that the value of P_p is negative.

10.11. FREE OSCILLATIONS IN A CIRCUIT WITH A PERIODICALLY VARYING CAPACITANCE

Let us set up an equation for a tuned circuit whose capacitance C varies according to

$$C(t) = C_0/(1 + m \cos \nu t) \qquad (10.72)$$

(L and r being constant) for the case where there is no external action:

$$L\frac{di}{dt}+ri+\frac{1}{C(t)}\int i(t)\,dt=0$$

Transforming from the current i to the charge q and taking into account expression (10.72), we get:

$$\frac{d^2q}{dt^2}+\frac{r}{L}\frac{dq}{dt}+\frac{(1+m\cos\nu t)}{LC_0}q=0 \qquad (10.73)$$

The quantity $1/LC_0=\omega_0^2$ determines the resonance frequency of the tuned circuit when there is no capacitance modulation, i.e., when $m=0$.

Thus, equation (10.73) can be written in the form

$$\frac{d^2q}{dt^2}+2\alpha_t\frac{dq}{dt}+\omega_0^2(1+m\cos\nu t)\,q=0 \qquad (10.74)$$

where $\alpha_t=r/2L$.

To reduce equation (10.74) to a canonical form, let us substitute

$$q=ye^{-\alpha_t t} \qquad (10.75)$$

this allowing the first derivative of q in (10.74) to be excluded.

Differentiating expression (10.75) twice to obtain

$$\frac{dq}{dt}=\left(\frac{dy}{dt}-\alpha_t y\right)e^{-\alpha_t t}$$

$$\frac{d^2q}{dt^2}=\left(\frac{d^2y}{dt^2}-2\alpha_t\frac{dy}{dt}+\alpha^2 y\right)e^{-\alpha_t t}$$

and substituting the obtained results into equation (10.74), we have

$$\frac{d^2y}{dt^2}+(-\alpha_t^2+\omega_0^2+m\omega_0^2\cos\nu t)\,y=\frac{d^2y}{dt^2}$$

$$+(\omega_{fr}^2+m\omega_0^2\cos\nu t)\,y=0 \qquad (10.76)$$

where $\omega_{fr}^2=\omega_0^2-\alpha_t^2$ is the square of the free-oscillation frequency of the tuned circuit (when there is no capacitance modulation).

Transforming to the dimensionless time

$$\tau=\nu t/2 \qquad (10.77)$$

and introducing the symbols

$$\delta=4\omega_{fr}^2/\nu^2,\quad \varepsilon=m4\omega_0^2/\nu^2 \qquad (10.78)$$

let us rewrite equation (10.76) in the form

$$y''+(\delta+\varepsilon\cos 2\tau)\,y=0 \qquad (10.79)$$

This equation is known as the *Mathieu equation*. The theory of the Mathieu equation is very well developed. Every value of the parameter ε has a corresponding succession of definite values of δ, for which the solutions of equation (10.79) are periodic Mathieu's functions which have a period of 2π with respect to the variable τ.

The relation between the values of ε and δ, for which periodic solutions of equation (10.79) exist, is shown in Fig. 10.25. For large values of ε, Mathieu's functions are rather complex. For $\varepsilon \ll 1$, these functions differ but slightly from harmonic oscillations. As $\varepsilon \to 0$, equations (10.76) and (10.79) transform respectively into the equations

$$y''(t) + \omega_{fr}^2 y(t) = 0 \tag{10.80}$$

$$y''(\tau) + \delta(0)\, y(\tau) = 0 \tag{10.80'}$$

Equation (10.80) has solutions in the form of the harmonic oscillations $\cos \omega_{fr} t$ and $\sin \omega_{fr} t$, while equation (10.80') has solutions $\cos \sqrt{\delta(0)}\,\tau$ and $\sin \sqrt{\delta(0)}\,\tau$. For the last two solutions to be periodic Mathieu's functions with a period of 2π relative to the variable τ, i.e., for them to have the form $\cos n\tau$ or $\sin n\tau$ (n is an integer), the value of $\delta(0)$ must be equal to the square of the integer n. From this it follows that for $\varepsilon = 0$ the curves shown in Fig. 10.25 must touch the abscissa axis at points $\delta(0) = n^2 = 1, 4, 9, \ldots$

For arbitrary values of ε and δ, the general solution of the Mathieu equation is

$$y = Ae^{\mu\tau}\varphi(\tau) + Be^{-\mu\tau}\varphi(-\tau) \tag{10.81}$$

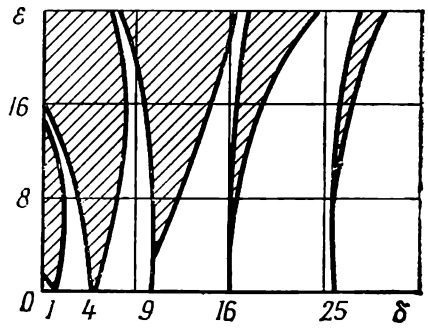


Fig. 10.25. Zones of stability (unshaded) and instability (shaded) of solutions of the Mathieu equation

where A and B are constants depending on the initial conditions, $\varphi(\tau)$ is a periodic Mathieu's function depending on the parameter ε and μ is an exponent depending on the parameters ε and δ of original equation (10.79).

Parametric systems used in r-f devices are characterized by a shallow capacitance (or inductance) modulation. As a rule, the modulation factor m is only a few per cent, so that the parameter ε, defined by formula (10.78), is small in comparison with unity. Therefore, in view of the above properties of Mathieu's functions, for $\varepsilon \ll 1$ we may write the general solution in the form

$$y \approx Ae^{\mu\tau}\cos(n\tau + \xi) + Be^{-\mu\tau}\cos(n\tau - \xi) = y_1(\tau) + y_2(\tau) \tag{10.82}$$

For further analysis, one should find the relation between the exponent μ and the parameters ε and δ, and also, determine the phase ξ.

To this end, let us substitute into original equation (10.79) one of the partial solutions, e.g., $y_1(\tau)$. To simplify the analysis, let us set $n = 1$, i.e., let us restrict ourselves to the examination of

the dependence of μ on ε and δ in the range of values of δ close to unity. Thus,

$$y_1(\tau) = Ae^{\mu\tau}\cos(\tau+\xi)$$

Substituting into (10.79) and grouping the terms containing $\cos\tau$ and $\sin\tau$, we get following equation*

$$\left\{\left[\frac{\varepsilon}{2}+(\mu^2+\delta-1)\right]\cos\xi-2\mu\sin\xi\right\}\cos\tau + \left\{\left[\frac{\varepsilon}{2}-(\mu^2+\delta-1)\right]\sin\xi-2\mu\cos\xi\right\}\sin\tau=0$$

Since this equality must hold for any values of τ, the coefficients multiplying $\cos\tau$ and $\sin\tau$ can be equated to zero.

Thus we obtain two equations for determining μ and ξ:

$$\left.\begin{aligned}\left[\frac{\varepsilon}{2}+(\mu^2+\delta-1)\right]\cos\xi&=2\mu\sin\xi\\ \left[\frac{\varepsilon}{2}-(\mu^2+\delta-1)\right]\sin\xi&=2\mu\cos\xi\end{aligned}\right\} \tag{10.83}$$

From the first equation it follows that

$$\tan\xi=\frac{\varepsilon/2+(\mu^2+\delta-1)}{2\mu} \tag{10.84}$$

and from the second

$$\tan\xi=\frac{2\mu}{\varepsilon/2-(\mu^2+\delta-1)} \tag{10.84'}$$

Equating the right-hand sides of expressions (10.84) and (10.84'), we obtain

$$(\varepsilon/2)^2-(\mu^2+\delta-1)^2=4\mu^2$$

or

$$\mu^4+2(\delta+1)\mu^2-[(\varepsilon/2)^2-(\delta-1)^2]=0$$

which yields

$$\mu^2=-(\delta+1)\pm\sqrt{(\delta+1)^2+(\varepsilon/2)^2-(\delta-1)^2} = -(\delta+1)\pm\sqrt{4\delta+(\varepsilon/2)^2}=-(\delta+1) \pm 2\sqrt{\delta}\sqrt{1+(\varepsilon/2)^2/4\delta} \tag{10.85}$$

From this expression it is obvious that μ^2 is a real number. Therefore, the exponent μ can either be a real number (positive or negative) or a purely imaginary number.

If μ is a positive or negative real number, the solution of the Mathieu equation is unstable, for one of the terms in expression (10.82) increases infinitely with an increase in τ. For imaginary values of μ, the solution is stable.

* The terms of the form $\cos 3\tau$ and $\sin 3\tau$ are omitted.

In the case of instability, the values of ε and δ are such that the points with the corresponding coordinates in Fig. 10.25 occur within one of the dashed zones. The curves shown in this figure are the boundaries separating the zones of stability and instability. The solid lines in Fig. 10.25 are the loci of points satisfying the condition $\mu = 0$. As mentioned above and also follows from expression (10.82), in this case the solutions of the Mathieu equation are periodic functions. The physical meaning of unstable solution is that for certain relations between the depth of capacitance modulation and the frequency of this modulation (ε and δ), in the tuned circuit there develop oscillations with an infinitely rising amplitude, no matter how weak the initial disturbances (e.g., thermal noise). The power source for these oscillations is the pumping oscillator acting on the capacitance.

Some unstable zones reach the abscissa axis at points $\delta = n^2$, where $n = 2\omega_{fr}/\nu = 1, 2, 3, \ldots$ This means that for $\delta = 1, 4, 9, \ldots$, i.e. for $\nu = 2\omega_{fr}$, $\nu = \omega_{fr}$, $\nu = \frac{2}{3}\omega_{fr}$, $\nu = \frac{1}{2}\omega_{fr}$, etc. the solutions of the Mathieu equation are unstable, no matter how shallow the modulation depth m ($\varepsilon \to 0$) of the parameter. For intermediate values of ν, in which case the parameter is acted upon out of step with the natural frequency of the tuned circuit, the lower the frequency ν the higher the value of ε is required for instability to occur.

In the region of low values of ε, expression (10.85) can be simplified [e.g., by taking the plus sign before the last term on the right-hand side of (10.85)]:

$$\mu^2 \approx -(\delta+1) + 2\sqrt{\delta}\left[1 + \frac{1}{2}\frac{1}{4\delta}\left(\frac{\varepsilon}{2}\right)^2\right]$$

$$= -(\sqrt{\delta}-1)^2 + \frac{1}{\sqrt{\delta}}\left(\frac{\varepsilon}{4}\right)^2$$

Thus,

$$\mu \approx \pm\sqrt{\frac{1}{\sqrt{\delta}}\left(\frac{\varepsilon}{4}\right)^2 - (\sqrt{\delta}-1)^2} \qquad (10.86)$$

Now, let us return to general solution (10.82) and transform from y to q [by formula (10.75)] and from τ to the dimensional time t [by formula (10.77)]:

$$q(t) = Ae^{\left(\frac{\nu}{2}\mu-\alpha_t\right)t}\cos\left(\frac{\nu}{2}t+\xi\right) + Be^{-\left(\frac{\nu}{2}\mu+\alpha_t\right)t}$$

$$\times\cos\left(\frac{\nu}{2}t-\xi\right) \qquad (10.87)$$

From this expression it is clear that for the system to be unstable (with respect to q) it is required that μ should be not only a real number but also should exceed in magnitude the ratio $2\alpha_t/\nu$.

Thus, the critical value of the exponent μ, corresponding to the boundary between the stable and unstable states of the system, is determined by the expression

$$|\mu_{cr}| = 2\alpha_t/\nu \tag{10.88}$$

Substituting this value of μ into formula (10.86), we find the critical value of ε:

$$\varepsilon_{cr} = 4\sqrt{[\mu_{cr}^2 + (\sqrt{\delta}-1)^2]\sqrt{\delta}} = 4\sqrt{4\alpha_t^2/\nu^2 + (\sqrt{\delta}-1)^2\sqrt{\delta}} \tag{10.89}$$

But according to (10.78)

$$\varepsilon_{cr} = m_{cr}(4\omega_0^2/\nu^2) \tag{10.90}$$

This gives m_{cr}, i.e., the critical value (on the verge of excitation) of the depth of capacitance modulation:

$$m_{cr} = \left(\frac{\nu}{2\omega_0}\right)^2 \varepsilon_{cr} = 4\left(\frac{\nu}{2\omega_0}\right)^2 \sqrt[4]{\delta}\sqrt{\frac{4\alpha_t^2}{\nu^2} + (\sqrt{\delta}-1)^2} \tag{10.91}$$

In the particular case $\nu = 2\omega_{fr} \approx 2\omega_0$ where $\delta = 1$, we have

$$m_{cr} = 4(2\alpha_t/\nu) = 4\alpha_t/\omega_0 = 2/Q = 2d \tag{10.92}$$

where $Q = 1/d$ is the quality factor of the tuned circuit.

This result agrees with condition (10.65) obtained from the examination of the steady-state operation of the parametric amplifier.

If the capacitance-modulation frequency ν is not exactly twice the frequency ω_{fr}, i.e., if there is a detuning

$$\Delta\omega = \nu/2 - \omega_{fr}$$

then the system becomes unstable when the value of m_{cr} exceeds $2/Q$.

In this case

$$\left.\begin{aligned} \delta &= \frac{4\omega_{fr}^2}{\nu^2} = \frac{4(\nu/2-\Delta\omega)^2}{\nu^2} \approx 1 - \frac{4\Delta\omega}{\nu} \approx 1 - \frac{2\Delta\omega}{\omega_{fr}} \\ (\sqrt{\delta}-1)^2 &= \left[\sqrt{1-\frac{2\Delta\omega}{\omega_{fr}}} - 1\right]^2 \approx \left[1 - \frac{\Delta\omega}{\omega_{fr}} - 1\right]^2 \approx \left(\frac{\Delta\omega}{\omega_{fr}}\right)^2 \\ m_{cr} &= 2\sqrt{d^2 + \left(\frac{2\Delta\omega}{\omega_{fr}}\right)^2} \end{aligned}\right\} \tag{10.93}$$

or

$$m_{cr}Q = 2\sqrt{1 + \left(\frac{2\Delta\omega}{\omega_{fr}}Q\right)^2} = 2\sqrt{1+a^2}, \quad a = (2\Delta\omega/\omega_{fr})Q \tag{10.93'}$$

The graph of $m_{cr}Q$ versus a is shown in Fig. 10.26 (solid curve).

The crosshatching denotes the excitation region (for $m > m_{cr}$), while the horizontal shading shows the region of unstable solutions of the

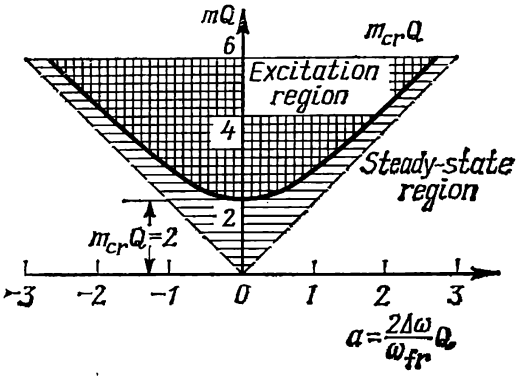


Fig. 10.26. Zones of stability and instability of solutions of the Mathieu equation in the vicinity of the point $\delta = 1$, $\varepsilon = 0$

Mathieu equation, which, however, become stable after being multiplied by $e^{-\alpha_t t}$. Finally, the unshaded region corresponds to stable solutions of the Mathieu equation and what is more, stable solutions of equation (10.74). Figure 10.26 is in fact an enlargement of the "tongue" in the vicinity of the point $\delta = 1$, $\varepsilon = 0$ in the plot of Fig. 10.25. All the parameters of a free oscillation in a parametric tuned circuit can easily be defined on the basis of the above analysis.

Let the periodic capacitance modulation according to (10.72) start at $t = 0$, the relations between the parameters m, ν, ω_{fr}, and Q being such that the system is unstable in the vicinity of the point $\delta = 1$, $\varepsilon = 0$.

By calculating $|\mu|$ by formula (10.86) and substituting into formula (10.84), we find the phase ξ. In particular, for $\delta = 1$ and $|\mu| = \varepsilon/4$, we have

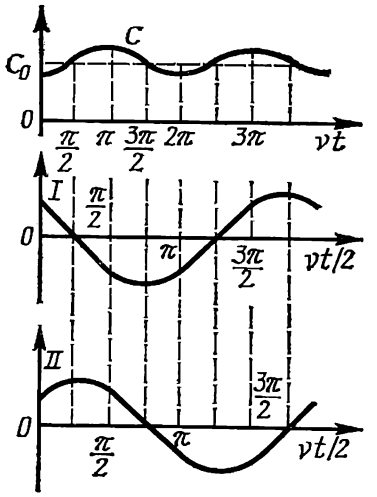


Fig. 10.27. Modulation of capacitance and the charge variation law providing (graph *I*) or not providing (graph *II*) for parametric excitation of an oscillation

$$\tan\xi = \{[\varepsilon/2 + (\varepsilon/4)^2]/2\,(\varepsilon/4)\} \approx 1, \quad \xi = 45°$$

Then, general solution (10.87) assumes the form

$$q(t) = A \exp\left[\left(\frac{\nu}{2}\mu - \alpha_t\right)t\right] \cos\left(\frac{\nu}{2}t + 45°\right)$$

$$+B\exp\left[-\left(\frac{\nu}{2}\mu+\alpha_t\right)t\right]\cos\left(\frac{\nu}{2}t-45^\circ\right)=q_1(t)+q_2(t) \qquad (10.94)$$

Figure 10.27 shows the graphs $C(t) = C_0/[1 + m\cos\nu t]$ and also $\cos(\nu t/2 + 45°)$ and $\cos(\nu t/2 - 45°)$. (In the given case $\nu/2 = \omega_{fr}$.) The last two graphs characterize the variation of the charges $q_1(t)$ and $q_2(t)$. From this figure it is clear that the capacitance decreases when q_1 passes through its maximum and q_2 passes through zero. This means that $q_1(t)$ is "correctly" phased with respect to $C(t)$ and the pumping causes the amplitude of q_1 to rise (according to $\exp[(\mu\nu/2 - \alpha_t)t]$), while $q_2(t)$ is phased incorrectly: the capacitance increases at the moments the charge passes through its maxima, this resulting in power takeoff from the tuned circuit and damping of the amplitude of q_2 according to $\exp[-(\mu\nu/2 + \alpha_t)t]$.

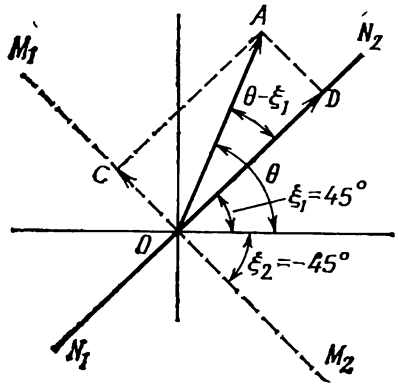


Fig. 10.28. Effect of initial conditions on parametric excitation

From the above discussion one can see how oscillations originate and rise in a parametric circuit. At the moment the tank circuit is energized (or at the moment the pumping oscillator is started) in the circuit there exist random noise oscillations due to thermal motion of charged particles.

These oscillations contain, among others, also a component of frequency $\nu/2$, however, the amplitude and phase of this component are random. Let us assume that at the initial moment this component has an amplitude OA and a phase θ (Fig. 10.28). Let us resolve the vector representing this oscillation along two mutually orthogonal directions: N_1N_2 and M_1M_2. The straight line N_1N_2 is drawn at an angle $\xi_1 = 45°$ to the abscissa axis, while the straight line M_1M_2 is drawn at an angle $\xi_2 = -45°$ to the same axis. The vector OD lying along the straight line N_1N_2 and equal to $OA\cos(\theta - \xi_1)$ represents the oscillation which is properly phased relative to the pumping voltage, while the vector OC lying along the straight line M_1M_2 represents the oscillation which begins to decay under the influence of the changing capacitance. We may therefore consider that the initial conditions for a parametric tuned circuit are determined by the component whose phase is in agreement with that of the pumping voltage. There are two such positions of the vector OD: $\xi_1 = 45°$ and $\xi_1 = 225°$ (for the particular case $\nu = 2\omega_{fr}$). If the vector OA is located to the right of the straight line M_1M_2, its projection onto the straight line N_1N_2 is positive, otherwise it is negative. This means that for a given phase of the pumping voltage,

the oscillation phase in the parametric tuned circuit can take on one of two fixed values differing by 180°.

In conclusion, it should be noted that the main results of this section can be extended to cover a parametric tuned circuit with a periodically varying inductance.

10.12. PARAMETRIC OSCILLATORS

From the preceding section it is evident that oscillations can be generated by periodically varying the parameter (capacitance or inductance) of one of the energy-storing elements of a tuned circuit.

This idea was first propounded by Soviet scientists L. I. Mandelshtam and N. D. Papaleksi who in 1931 developed the theory of parametric excitation and proved it experimentally on a model of a tank circuit in which the inductance or capacitance was modulated by a mechanical means (for instance, by rotating the rotor of a variometer).

The principle of parametric excitation of oscillations is now widely used in special oscillators—so-called *parametrons* — employed in various digital circuits. This is explained by the fact that in these devices a binary digit can be presented by selecting a proper phase angle. Since the phase φ or $\varphi + \pi$ depends on the initial conditions, by specifying the initial phase angle by a signal when the oscillator is switched on, it is possible to obtain two stable states of the oscillator corresponding to the two digits of the binary code (for example, the phase φ corresponds to zero, and the phase $\varphi + \pi$ to unity).

In a capacitive parametron (Fig. 10.29*a*) two semiconductor diodes are used as a variable capacitance, while the primary winding of a high-frequency transformer serves as an inductance. The pumping voltage $e_p(t)$ having a frequency ω_p twice as high as the resonance frequency of the tuned circuit, is applied to the diodes in the same phase, so that the capacitances of the diodes increase or decrease simultaneously, and at the same time the frequency ω_p is rejected and does not reach the output. On the other hand, due to the symmetry of the system, the frequency $\omega_p/2$ excited in the tuned circuit is also rejected and does not reach the pumping circuit. The position of the operating point on the curves of the *p-n* junctions is determined by a d-c bias voltage.

In an inductive parametron (Fig. 10.29*b*) the tuned circuit consists of a fixed capacitor and coils L_t on ferrite cores whose magnetic permeability changes periodically during the passage of the pumping current $i_p(t)$ through the coils L_p. The initial position of the operating point on the nonlinear inductance characteristic is determined by a d-c current passing through the coils L_p. The opposite connection of the coils L_p on the two cores rejects the direct passage of the oscillation with a frequency ω_p to the output, as well as the

oscillations with a frequency $\omega_p/2$ from the tank to the pumping circuit.

The conditions necessary for the origination and rise of oscillations in a linear parametric tuned circuit were discussed in detail in the preceding section. However, to determine the stationary amplitude in a parametric oscillator, we must consider the nonlinearity which inevitably manifests itself in the process of the amplitude rise and on which the mechanism of amplitude limiting depends.

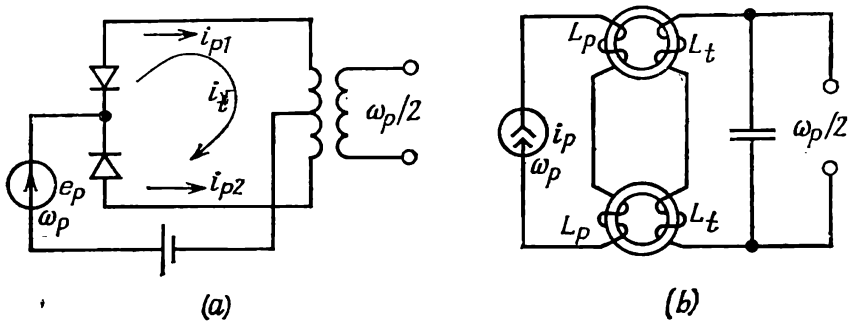


Fig. 10.29. (*a*) Capacitive and (*b*) inductive parametrons

In the case of a parametric oscillator with a mechanical device for modulating the energy-storing element, the amplitude rise is limited by the power of the pumping device. Besides, amplitude limiting can result from the fact that the oscillation amplitude may get onto the nonlinear sections of the characteristics of the nonlinear capacitance or inductance. In this case, the average values of $C(t)$ or $L(t)$ and consequently, the mean value of the resonance frequency of the tuned circuit are changed. The detuning of the tank circuit relative to the frequency $\omega_p/2$ affects the conditions of the pumping energy conversion and leads to amplitude limiting.

It should be noted that the term "oscillator" is only conventional when applied to a parametron. In contrast to any electronic self-oscillating system or any externally excited oscillator, in which d-c energy is converted into a-c energy, it is the pumping oscillator that is the *primary energy source* in a parametron.

The mission of a parametron used as a relay with two stable states is to be not a source of oscillations, but a memory device storing the phase of a signal.

In this connection, its fast response is of special importance, since it determines the response time of the devices employing parametrons. The rate of amplitude rise must be as high as possible for every start-up of the parametron.

Since, in accordance with formula (10.94), the oscillation amplitude in the tuned circuit increases as

$$A(t) = A_0 e^{[(\nu/2)\mu - \alpha_t]t}$$

where A_0 is the initial amplitude (i.e. the amplitude of the signal whose phase is to be stored), the rate of amplitude rise at the starting moment is

$$v_0 = \frac{dA(t)}{dt}\Big|_{t=0} = \left(\frac{\nu}{2}\mu - \alpha_t\right) A_0$$

Taking into account that, according to expressions (10.86) and (10.78), for $\nu/2 = \omega_0$ the parameter $\mu = \varepsilon/4 = m$, and also that $\alpha_t/\omega_0 = d/2 = 1/2Q$, we have

$$v_0 = \left(\frac{\nu}{2\omega_0}\mu - \frac{\alpha_t}{\omega_0}\right)\omega_0 A_0 = \left(m - \frac{d}{2}\right)\omega_0 A_0$$

The possibility of raising the parameter m and the amplitude A_0 is rather limited. Consequently, the basic way of improving the response time of the system is to raise the frequency ω_0.

The operating frequencies of parametrons are being continuously raised, and new electron and other devices are being developed for parametric operation, allowing one to construct parametrons working in the super-high frequency range.

Chapter 11

EFFECT OF RANDOM OSCILLATIONS ON NONLINEAR AND PARAMETRIC CIRCUITS

11.1. GENERAL

When analyzing the noise immunity of radio systems, one most frequently has to consider a linear sum of a useful signal $s(t)$ and a noise signal $n(t)$:

$$y(t) = s(t) + n(t) \tag{11.1}$$

In this case, noise is called *additive*, while $y(t)$ is an additive mixture of signal and noise. Additive noise is exemplified by shot and thermal noise considered in Ch. 7, which develop in electron devices and electric circuits regardless of the signals acting therein.

However, when a signal is transmitted through an actual communication channel, in addition to additive noise, there are other factors which distort this signal. Such factors include parasitic time variations of the parameters of the circuits or any other elements of the communication channel. In a simplest case where these variations have the character of amplitude modulation, the signal at the channel output can be presented in the form

$$s_{out}(t) = K(t)\, s(t) + n(t)$$

In this expression, $n(t)$, as in (11.1), is additive noise, while $K(t)$ is a coefficient characterizing *multiplicative* noise. Under actual conditions, the formation mechanism of multiplicative noise is more complex and cannot always be reduced to simple multiplication of signal and noise. In spite of this, by multiplicative noise is usually meant such noise as results from undesirable changes in the parameters of the linear system through which the signal is being transmitted.

The first sections of this chapter deal with the action of normal, mainly narrow-band noise on nonlinear devices such as amplitude and frequency detectors, nonlinear amplifiers and amplitude limiters. Sections 11.8 and 11.9 are devoted to the study of the action of random processes on parametric circuits and of the effect of multiplicative noise on the transmission of signals.

11.2. CONVERSION OF A RANDOM PROCESS IN LAG-FREE NONLINEAR CIRCUITS

An actual nonlinear device is a combination of lag-free nonlinear elements with time-lag linear electric circuits. This complicates a lot the determination of statistical characteristics of the signal and

noise at the output of the whole system. For linear circuits, one can easily find the autocorrelation (or spectral) function, but it is very difficult to determine the distribution law. On the contrary, in nonlinear but lag-free elements the main difficulty consists in finding the autocorrelation function. Therefore, there are no general methods for analyzing the conversion of random processes in nonlinear systems. One has to restrict oneself to some particular problems of practical interest, that are capable of being solved, and also use various idealizations of characteristics of the model of the device under study.

Let a nonlinear element be acted upon by a random oscillation (voltage, current) with a given probability density $p(x)$. It is necessary to find the probability density $p(y)$ of the output quantity y. The relation between y and x is determined by the nonlinear dependence $y = f(x)$ having the meaning, for example, of the current-voltage characteristic of an electron, semiconductor or some other active element.

If $f(x)$ defines the single-valued relation between x and y at each given moment irrespective of the values of x at the preceding moments of time (a lag-free element), the probability density $p(y)$ is found from the obvious expression

$$p(y)\,dy = p(x)\,dx \tag{11.2}$$

whence, taking into account that $p(x)$ and $p(y)$ are non-negative,

$$p(y) = p(x)/|\,dy/dx\,| \tag{11.3}$$

If the inverse function $x = \varphi(y)$ is many-valued, we have

$$p(y) = \left[\frac{p(x)}{|\,dy/dx\,|}\right]_{x=x_1} + \left[\frac{p(x)}{|\,dy/dx\,|}\right]_{x=x_2} + \ldots \tag{11.4}$$

where $x_1, x_2, \ldots$ are the values of the input quantity x corresponding to the given value of y.

If the characteristic $y = f(x)$ is constant within some range of x, expression (11.3) should be supplemented with a term with a delta function, accounting for the integral probability of residence of x below (or above) a definite level.

Let us illustrate the determination of $p(y)$ with some practical examples.

1. The effect of a normally distributed random process $x(t)$ on an element with a symmetrical quadratic characteristic (Fig. 11.1). The current-voltage characteristic shown in Fig. 11.1 can, for example, be realized by means of push-pull connection of two diodes having a quadratic characteristic near zero (Fig. 11.2).

With the voltage polarity being as shown in Fig. 11.2, a current equal to a_2x^2 flows through the diode D_1, with the opposite polarity, it flows through the diode D_2.

Setting $y = a_2x^2$, $dy/dx = 2a_2x$ and taking into account the fact that two values of x, namely, $x_1 = +\sqrt{y/a_2}$ and $x_2 = -\sqrt{y/a_2}$,

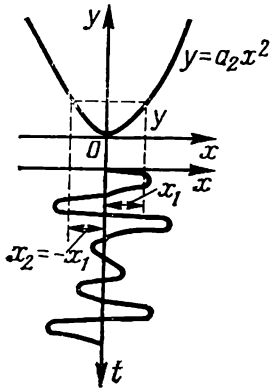


Fig. 11.1. Effect of a random process on a nonlinear element with a quadratic characteristic

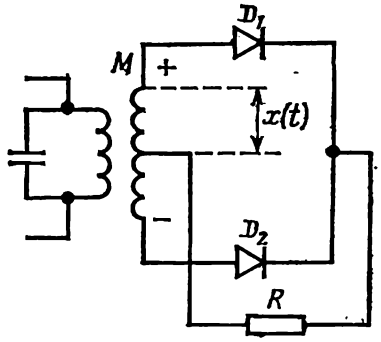


Fig. 11.2. ush-pull connection of diodes

correspond to some fixed value of y, we find by formula (11.4)

$$p(y) = \begin{cases} p(+\sqrt{y/a_2})/2a_2\sqrt{y/a_2} \\ \quad +p(-\sqrt{y/a_2})/2a_2\sqrt{y/a_2} & \text{for } y \geqslant 0 \\ 0 & \text{for } y < 0 \end{cases} \qquad (11.5)$$

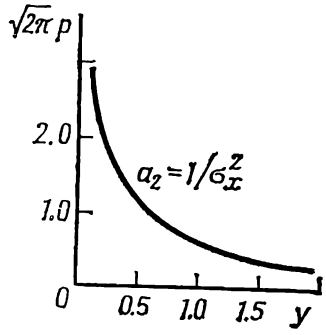


Fig. 11.3. Probability density of the current in a circuit with a quadratic current-voltage characteristic, which is acted upon by a normal random process

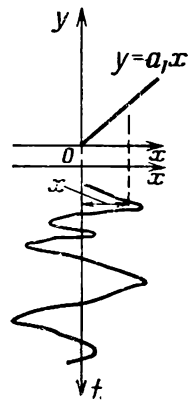


Fig. 11.4. Effect of a random function on a half-wave detector

Substituting $x_{1,2}^2 = y/a_2$ into the expression for the probability density $p(x)$

$$p(x_{1,2}) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-x_{1,2}^2/2\sigma_x^2} = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-y/2a_2\sigma_x^2}$$

we finally obtain

$$p(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,\sigma_x \sqrt{a_2}} \dfrac{e^{-y/2a_2\sigma_x^2}}{\sqrt{y}} & \text{for} \quad y \geqslant 0 \\ 0 & \text{for} \quad y < 0 \end{cases} \tag{11.6}$$

The graph of this distribution is shown in Fig. 11.3 (for $a_2 = 1/\sigma_x^2$).

2. The effect of a normally distributed process on a half-wave detector with a piecewise-linear characteristic (Fig. 11.4).

In this case

$$y = \begin{cases} a_1 x & \text{for} \quad x \geqslant 0 \\ 0 & \text{for} \quad x < 0 \end{cases}$$

It is clear that, in accordance with (11.3),

$$p(y) = \frac{p(x = y/a_1)}{a_1} = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,a_1\sigma_x} e^{-y^2/2\sigma_x^2 a_1^2} & \text{for} \quad y > 0 \\ 0 & \text{for} \quad y < 0 \end{cases}$$

Special attention should be paid to the behaviour of the function $p(y)$ at the point $y = 0$. Since $y = 0$ for any negative values of x, the probability $P(y = 0)$ is equal to the probability that $x \leqslant 0$. But the probability $P(x \leqslant 0) = 1/2$. From this it follows that the probability density $p(y = 0) = \infty$.

This fact can be taken account of by writing the expression for $p(y)$ in the form

$$p(y) = \begin{cases} \dfrac{1}{2}\delta(y) + \dfrac{1}{\sqrt{2\pi}\,a_1\sigma_x} e^{-y^2/2a_1^2\sigma_x^2} & \text{for} \quad y \geqslant 0 \\ 0 & \text{for} \quad y < 0 \end{cases} \tag{11.7}$$

The term $\delta(y)/2$ is zero everywhere except for the point $y = 0$ where it goes into infinity. But when being integrated over y, this term yields 1/2. The graphs of $p(x)$ and $p(y)$ are shown in Fig. 11.5.

3. The effect of a normally distributed process on a limiter (Fig. 11.6).

By analogy with the preceding case, one can easily set up the expression

$$p(y) = \begin{cases} (1/2)\,\delta(y) + \dfrac{1}{a_1\sqrt{2\pi}\,\sigma_x} e^{-y^2/2a_1^2\sigma_x^2} \\ + P(x > x_0)\,\delta(y - y_0) & \text{for} \quad 0 \leqslant y \leqslant y_0 \\ 0 & \text{for} \quad y < 0 \text{ and } y > y_0 \end{cases} \tag{11.8}$$

The graphs of the distributions of x and y are shown in Fig. 11.7. These examples are sufficient for one to understand the method of

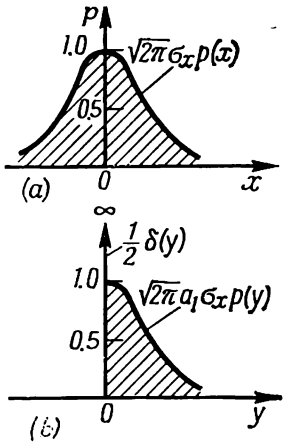


Fig. 11.5. Probability density of a random process at (*a*) the input and (*b*) the output of a half-wave detector

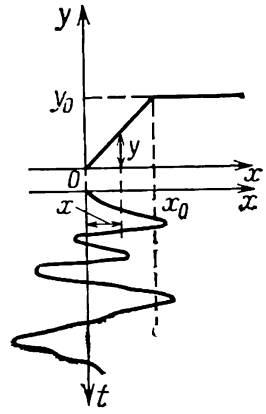


Fig. 11.6. Effect of a random function on a limiter

finding the probability density of a random variable at the output of a lag-free nonlinear element with any current-voltage characteristic. The simplicity of this method is due to the fact that the effect of the output (time-lag) circuits on the operation of the nonlinear element in question is not taken into account.

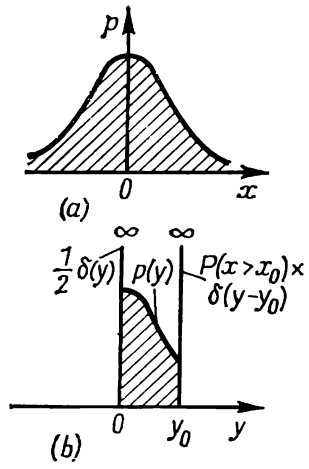


Fig. 11.7. Probability density of a random process at (*a*) the input and (*b*) the output of a limiter

11.3. CONVERSION OF ENERGY SPECTRUM IN A LAG-FREE NONLINEAR ELEMENT

It is impossible to determine the energy spectrum of the output process directly from the known spectrum at the input of a nonlinear element. The only way is to determine the autocorrelation function and then apply the Fourier transformation.

If a stationary process $x(t)$ acts at the input of a nonlinear element having a characteristic $y = f(x)$, the autocorrelation function at the output can be represented in the form

$$B_y(\tau) = \langle y_t y_{t+\tau} \rangle = \langle f(x_t)\, f(x_{t+\tau}) \rangle \qquad (11.9)$$

where x_t and $x_{t+\tau}$ are the values of $x(t)$ at the moments t and $t + \tau$, and y_t and $y_{t+\tau}$ are the corresponding values of y at the output of the nonlinear element.

To average the product $f(x_t) f(x_{t+\tau})$, one must know the bivariate probability density $p(x_t, x_{t+\tau})$ of the input process. If this probability density is known, the autocorrelation function $B_y(\tau)$ can be represented in the form

$$B_y(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1) f(x_2)\, p(x_1, x_2)\, dx_1\, dx_2 \qquad (11.10)$$

where x_1 and x_2 stand for x_t and $x_{t+\tau}$, respectively.

In some practically important problems this integral cannot be calculated. In this connection, one often has to use various indirect methods, one of which will be described later in the text.

As an example of a problem of practical interest, that can be solved directly, let us consider the effect of a stationary normal process $x(t)$ on a nonlinear element with a quadratic characteristic $y = a_2 x^2$ (see example 1 of the preceding section).

The bivariate probability density of the process $x(t)$ is*

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_x^2 \sqrt{1-R^2}} \exp\left[-\frac{x_1^2 + x_2^2 - 2Rx_1x_2}{2\sigma_x^2(1-R^2)}\right] \qquad (11.11)$$

where R is the correlation coefficient of x_1 and x_2, i.e., $R = R_x(\tau)$.

Substituting expression (11.11) and also $f(x) = a_2 x^2$ into (11.10), we obtain

$$B_y(\tau) = \frac{a_2^2}{2\pi\sigma_x^2 \sqrt{1-R^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 x_2^2 \exp\left[-\frac{x_1^2 + x_2^2 - 2Rx_1x_2}{2\sigma_x^2(1-R^2)}\right] dx_1\, dx_2$$

$$= \frac{a_2^2}{2\pi\sigma_x^2 \sqrt{1-R^2}} \int_{-\infty}^{\infty} x_1^2 \exp\left[-\frac{x_1^2}{2\sigma_x^2(1-R^2)}\right]$$

$$\times \left\{\int_{-\infty}^{\infty} x_2^2 \exp\left[-\frac{x_2^2 - 2Rx_1x_2}{2\sigma_x^2(1-R^2)}\right] dx_2\right\} dx_1 \qquad (11.12)$$

The integral in the braces can easily be calculated, completing the square in the numerator $x_2^2 - 2Rx_1x_2$ to obtain $(x_2 - Rx_1)^2 - R^2x_1^2$ and replacing the variable $x_2 - Rx_1 = z$:

$$\exp\left[\frac{R^2x_1^2}{2\sigma_x^2(1-R^2)}\right] \int_{-\infty}^{\infty} (z^2 + 2Rx_1z + R^2x_1^2)\, e^{-\frac{z^2}{2\sigma_x^2(1-R^2)}} dz$$

$$= \exp\left[\frac{R^2x_1^2}{2\sigma_x^2(1-R^2)}\right] \left[\sqrt{2\pi}\,\sigma_x^3 (1-R^2)^{3/2}\right.$$

$$\left. + 0 + \sqrt{2\pi}\,\sigma_x \sqrt{1-R^2}\, R^2x_1^2\right]$$

* See, for example, [1].

Substituting this result into (11.12), we obtain

$$B_y(\tau) = \frac{a_2^2}{\sqrt{2\pi}\,\sigma_x}\left[\sigma_x^2(1-R^2)\int_{-\infty}^{\infty} x_1^2 e^{-x_1^2/2\sigma_x^2}\,dx_1 + R^2\int_{-\infty}^{\infty} x_1^4 e^{-x_1^2/2\sigma_x^2}\,dx_1\right]$$

After that we determine

$$\int_{-\infty}^{\infty} x_1^2 e^{-x_1^2/2\sigma_x^2}\,dx_1 = \sqrt{2\pi}\,\sigma_x^3$$

and

$$\int_{-\infty}^{\infty} x_1^4 e^{-x_1^2/2\sigma_x^2}\,dx_1 = \sqrt{2\pi}\,3\sigma_x^5$$

Thus,

$$B_y(\tau) = a_2^2\sigma_x^4[(1-R^2)+3R^2] = a_2^2\sigma_x^4 + 2a_2^2\sigma_x^4 R_x^2(\tau)$$
$$= a_2^2\sigma_x^4 + 2a_2^2 B_x^2(\tau) \qquad (11.13)$$

Here we have used the well-known relation $R_x(\tau) = B_x(\tau)/\sigma_x^2$ (for $\langle x\rangle = 0$).

Of special interest is the effect of a *narrow-band* random process on a nonlinear element (detection problem).

Representing the autocorrelation function of a narrow-band proces in the form (4.76) and taking into account that

$$B_x^2(\tau) = \sigma_x^4 R_0^2(\tau)\left[\frac{1}{2} + \frac{1}{2}\cos 2\omega_0\tau\right], \qquad (11.14)$$

where R_0 is the envelope of the autocorrelation function of the narrow-band process, we write expression (11.13) in the final form

$$B_y(\tau) = a_2^2\sigma_x^4 + a_2^2\sigma_x^4 R_0^2(\tau) + a_2^2\sigma_x^4 R_0^2(\tau)\cos 2\omega_0\tau \qquad (11.15)$$

After that, using the Fourier transformation we obtain the general expression for the energy spectrum of the process at the output of a quadratic element (with a normal process at the input):

$$W_y(\omega) = a_2^2\sigma_x^4 2\pi\delta(\omega) + a_2^2\sigma_x^4\int_{-\infty}^{\infty} R_0^2(\tau)\,e^{-i\omega\tau}\,d\tau$$
$$+ a_2^2\sigma_x^4\int_{-\infty}^{\infty} R_0^2(\tau)\cos(2\omega_0\tau)\,e^{-i\omega\tau}\,d\tau$$
$$= W_{y0}^{\text{дис}}(\omega) + W_{l\text{-}f}(\omega) + W_{h\text{-}f}(\omega) \qquad (11.16)$$

The first term (discrete) corresponds to the d-c component of the output oscillation, the second term corresponds to the low-frequency

fluctuation component whose spectrum adjoins the zero frequency, and the third term corresponds to the high-frequency fluctuation component with a spectrum grouping near the frequency $2\omega_0$.

11.4. EFFECT OF NARROW-BAND NOISE ON AN AMPLITUDE DETECTOR

An amplitude detector comprising a diode and a low-pass filter (*R-C* circuit) is a combination of a lag-free nonlinear element and a time-lag linear network.

Let us divide the given device into two parts: (1) a nonlinear element and (2) a low-pass filter.

The methods discussed in the preceding sections, as well as some other special methods, make it possible, in principle, to find the distribution law and the autocorrelation function of noise first at the output of the nonlinear element (diode) and then at the output of the filter. In the general case, these investigations require fairly cumbersome calculations. The problem can be made much easier to solve by using some simplifications based on the operating principles of actual devices.

First, let us consider the so-called "linear" detection, i.e., detection of a high-frequency oscillation with a sufficiently high amplitude. By this oscillation here is meant normal noise (when the signal is absent) formed by the selective circuits at the detector input. As in the case of detection of a useful amplitude-modulated oscillation, we may consider that the output voltage of the "linear" detector reproduces the envelope of the high-frequency oscillation, namely, the noise envelope. Therefore, in the case of linear detection, there is no need in considering separately the statistical characteristics of the diode current and the output voltage of the *R-C* circuit. The voltage $v_{out}(t)$ existing across this circuit can be taken to be equal to the noise envelope $V(t)$ at the input of the detector (i.e., the transfer factor of the detector can be considered to be equal to unity). In such an approach, the statistical characteristics of noise at the output of the detector completely coincide with the characteristics of the envelope $A(t)$, given in Sec. 4.6. Thus, we conclude that the noise voltage at the output of the linear detector has the Rayleigh distribution

$$p(v_{out}) = p(A) = \frac{v_{out}}{\sigma_x^2} e^{-v_{out}^2/2\sigma_x^2}, \qquad (11.17)$$

$$0 < v_{out} < \infty$$

Using formulas (4.71) and (4.72), we find the mean value (the d-c component) of the noise voltage

$$V_0 = \langle v_{out}(t) \rangle = \langle A(t) \rangle = \sqrt{\pi/2}\,\sigma_x = 1.26\sigma_x \qquad (11.18)$$

and the mean square of the voltage

$$\langle v_{out}^2(t)\rangle = 2\sigma_x^2 \tag{11.19}$$

From this it follows that the variance of noise at the output of the linear detector is

$$\sigma_{out}^2 = \langle v_{out}^2\rangle - V_0^2 = 2\sigma_x^2 - \frac{\pi}{2}\sigma_x^2 = 0.43\sigma_x^2 \tag{11.20}$$

Thus, the basic parameters of the output noise (the d-c component V_0 and the variance σ_{out}^2) are simply expressed through the variance σ_x^2 of the high-frequency noise at the detector input.

The autocorrelation function and the energy spectrum of the output noise can easily be found by formulas (4.77) and (4.78).

As an example, let us consider the effect on the linear detector of normal noise $x(t)$ whose energy spectrum is defined by the expression

$$W_x(\omega) = N_0\left[e^{-\alpha(\omega-\omega_0)^2} + e^{-\alpha(\omega+\omega_0)^2}\right] \tag{11.21}$$

and the autocorrelation function, according to (4.39) and in view of [6] [formulas (3.896.3) and (3.896.4)], is

$$B_x(\tau) = N_0\frac{1}{2\pi}\left[\int_{-\infty}^{\infty} e^{-\alpha(\omega-\omega_0)^2}e^{i\omega\tau}\,d\omega + \int_{-\infty}^{\infty} e^{-\alpha(\omega+\omega_0)^2}e^{i\omega\tau}\,d\omega\right]$$

$$= \frac{N_0}{\sqrt{\pi\alpha}}e^{-\tau^2/4\alpha}\cos\omega_0\tau = \sigma_x^2 e^{-\tau^2/4\alpha}\cos\omega_0\tau \tag{11.22}$$

Then

$$R_0(\tau) = e^{-\tau^2/4\alpha} \tag{11.23}$$

and, in accordance with (4.78),

$$W_{out}(\Omega) = \frac{\pi\sigma_x^2}{2}\left[2\pi\delta(\Omega) + \frac{1}{4}\int_{-\infty}^{\infty} e^{-\tau^2/2\alpha}\,e^{-i\Omega\tau}\,d\tau\right]$$

$$= \frac{\pi\sigma_x^2}{2}\left[2\pi\delta(\Omega) + \frac{1}{4}\sqrt{2\pi\alpha}\,e^{-\alpha\Omega^2/2}\right] \tag{11.24}$$

The term with the delta function corresponds to the d-c voltage component at the detector output.

The graph of $W_{out}(\Omega)$ is shown in Fig. 11.8*b*. The width of this spectrum exceeds that of the spectrum $W_x(\omega)$ at the detector input (Fig. 11.8*a*) by a factor of $\sqrt{2}$.

The linear amplitude detector reproduces the envelope of a narrow-band oscillation regardless of the specific features of its spectrum structure. The result obtained indicates that the envelope of each of

the realizations of the noise in question (at the detector input) has a wider spectrum compared to the frequency band of the realization itself. At first glance it might seem strange, because it is well known that for a modulated oscillation the width of the envelope spectrum is either the same as that of the spectrum of the oscillation (in AM) or is narrower than that (in FM). This apparent discrepancy is easily settled if one takes into account the fact that the oscillations of the lower and upper side frequencies in modulation are fully correlated. For example, it is sufficient to disturb the symmetry of the amplitudes or phases of the side-frequency oscillations in amplitude modulation for the sum of the three oscillations of frequencies ω_0, $\omega_0 + \Omega$ and $\omega_0 - \Omega$ to turn into an oscillation whose envelope contains, in addition to the frequency Ω, also frequencies 2Ω, 3Ω, etc. In this case, the amplitude detector produces an output oscillation whose spectrum is wider than the frequency band of the high-frequency input oscillation. On the other hand, in the spectrum of noise there is no correlation (symmetry) at all between the spectral components whose frequencies are located to the left and right of the centre frequency ω_0. Naturally, the envelope of each realization of noise has a wider spectrum than a modulated wave having the same spectrum width. Correspondingly, the average spectrum width of the envelope, i.e., its energy spectrum, is also increased.

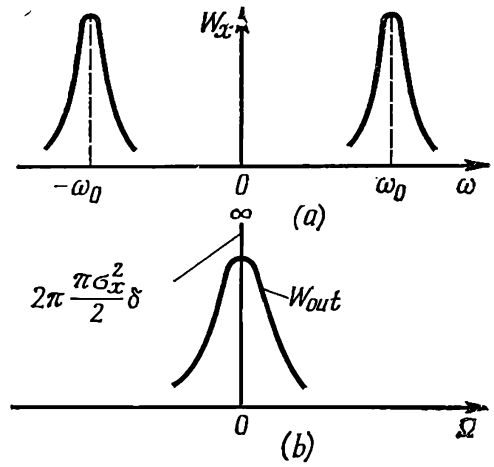


Fig. 11.8. Energy spectrum of a random process at (*a*) the input and (*b*) the output of an amplitude detector

Now let us consider the effect of normal noise on a square-law detector. In this case, the output voltage of the detector (taking into consideration the rejection of the high-frequency component of noise) by analogy with expression (8.55) can be written in the form

$$v_{out}(t) = KA^2(t)/2 \tag{11.25}$$

where K is the coefficient accounting for the parameter a_2 of the current-voltage characteristic of the diode and the value of the load resistor at the output of the detector.

Using formula (11.3) in which by $p(x)$ one should mean the probability density of the envelope $A(t)$, we find the distribution law

of the noise voltage at the output of the square-law detector:

$$p(v_{out}) = p\left(K\frac{A^2}{2}\right) = \frac{p(A)}{\left|\frac{dv_{out}}{dA}\right|}$$

$$= \frac{A}{\sigma_x^2} e^{-A^2/2\sigma_x^2} \frac{1}{KA} = \frac{1}{K\sigma_x^2} e^{-v_{out}/K\sigma_x^2} \quad (11.26)$$

Thus, when a normal narrow-band noise acts on a square-law detector with a low-pass filter, the output noise of the entire device has an *exponential distribution.*

Now let us calculate the mean value of the output voltage:

$$\langle v_{out}(t)\rangle = \int_0^\infty v_{out} p(v_{out})\, dv_{out} = \frac{1}{K\sigma_x^2}\int_0^\infty v_{out} e^{-v_{out}/K\sigma_x^2}\, dv_{out} = K\sigma_x^2 \quad (11.27)$$

and also the mean square of the output voltage:

$$\langle v_{out}^2(t)\rangle = \int_0^\infty v_{out}^2 p(v_{out})\, dv_{out} =$$

$$= \frac{1}{K\sigma_x^2}\int_0^\infty v_{out}^2 e^{-v_{out}/K\sigma_x^2}\, dv_{out} = 2K^2\sigma_x^4 \quad (11.28)$$

From this it follows that the variance of noise at the output is

$$\sigma_{out}^2 = \langle v_{out}^2(t)\rangle - \langle v_{out}(t)\rangle^2 = 2K^2\sigma_x^4 - K^2\sigma_x^4 = K^2\sigma_x^4 \quad (11.29)$$

For complete description of the properties of noise at the output of a square-law detector, it only remains to calculate its autocorrelation function and the energy spectrum. This can be accomplished by means of formulas (11.15) and (11.16). The second term in expression (11.15) defines the autocorrelation function sought for, and the second term in expression (11.16) is the energy spectrum corresponding to this function.

For $R_0(\tau) = e^{-\tau^2/4\alpha}$ (see the preceding example), we have

$$W_{out}(\Omega) = W_{y0}(\Omega) + W_{l\text{-}f}(\Omega) = a_2^2\sigma_x^4\left[2\pi\delta(\Omega) + \int_{-\infty}^{\infty} e^{-\tau^2/2\alpha} e^{-i\Omega\tau}\, d\tau\right]$$

$$= a_2^2\sigma_x^4\left[2\pi\delta(\Omega) + \sqrt{2\pi\alpha}\, e^{-\alpha\Omega^2/2}\right] \quad (11.30)$$

The graphs of the functions $W_x(\omega)$ and $W_{out}(\Omega)$ have the same shape as the graphs in Fig. 11.8, the only difference being in the scale of ordinates due to the difference in the constant coefficients [$a^2\sigma_x^4$ instead of $\pi\sigma_x^2/2$ before the brackets in (11.24) and 1 instead of 1/4 before the second term].

11.5. JOINT EFFECT OF A HARMONIC OSCILLATION AND NORMAL NOISE ON AN AMPLITUDE DETECTOR

When a narrow-band noise $x(t) = A(t) \cos[\omega_0 t + \theta(t)]$ is superimposed on a signal $s(t) = E \cos \omega_0 t$, resultant oscillation can be written in the form

$$v(t) = s(t) + x(t) = E \cos \omega_0 t + A(t) \cos[\omega_0 t + \theta(t)]$$

$$= [E + A(t) \cos \theta] \cos \omega_0 t - A(t) \sin \theta(t) \sin \omega_0 t$$

$$= V(t) \cos[\omega_0 t + \xi(t)] \quad (11.31)$$

where the envelope $V(t)$ and phase $\xi(t)$, by analogy with (8.43) and (8.44), are defined by the expressions

$$V(t) = \sqrt{E^2 + A^2(t) + 2EA(t) \cos \theta(t)} \quad (11.32)$$

$$\xi(t) = \arctan \frac{A(t) \sin \theta(t)}{E + A(t) \cos \theta(t)} \quad (11.33)$$

When analyzing the action of the oscillation on an amplitude detector, the statistical characteristics of the phase $\xi(t)$ can be disregarded (this question, as applied to a frequency detector, will be discussed in the next section). Of primary importance is the probability density $p(V)$ of the envelope V determined by the formula [1]

$$p(V) = \frac{V}{\sigma_x^2} e^{-(V^2+A^2)/2\sigma_x^2} I_0 \left(\frac{E}{\sigma_x} \frac{V}{\sigma_x}\right) \quad (11.34)$$

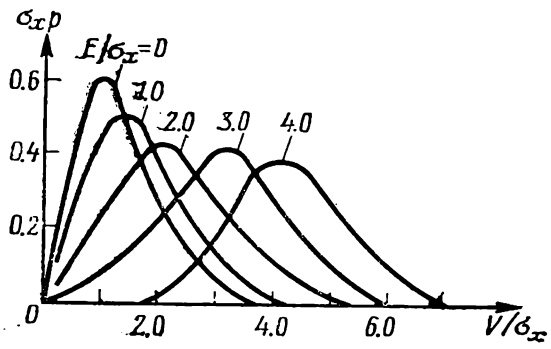


Fig. 11.9. Generalized Rayleigh probability density

where I_0 is a Bessel function (modified) of complex argument.

The function determined by formula (11.34) is called the *generalized Rayleigh function*. The graphs of the function $p(V)$ for several values of E/σ_x are shown in Fig. 11.9. For $E/\sigma_x = 0$ (no signal), expression (11.34) transforms into (4.70). In the other extreme case, where the amplitude E of the signal is very high compared to σ_x, the curve $p(V)$ is close to a Gaussian curve with the variance σ_x^2 and the mean equal to E.

Let us first consider linear detection. We shall presume that the voltage at the detector output coincides with the envelope of the r-f voltage at the input. In this case, using formula (11.34), we find the

d-c component of the voltage at the detector output

$$V_0 = \langle V \rangle = \int_0^\infty V p(V)\, dV = \frac{1}{\sigma_x^2} e^{-E^2/2\sigma_x^2} \int_0^\infty V^2 e^{-V^2/2\sigma_x^2} I_0\left(\frac{E}{\sigma_x}\frac{V}{\sigma_x}\right) dV$$

and the mean square of the voltage

$$\langle V^2(t) \rangle = \int_0^\infty V^2 p(V)\, dV = \frac{1}{\sigma_x^2} e^{-E^2/2\sigma_x^2} \int_0^\infty V^3 e^{-V^2/2\sigma_x^2} I_0\left(\frac{E}{\sigma_x}\frac{V}{\sigma_x}\right) dV$$

Calculating the integrals, [5], we obtain the following expressions

$$\left.\begin{aligned} V_0 &= \sigma_x \sqrt{\frac{\pi}{2}} \left\{ I_0(E^2/4\sigma_x^2) + \frac{E^2}{2\sigma_x^2} [I_0(E^2/4\sigma_x^2) \right. \\ &\quad \left. + I_1(E^2/4\sigma_x^2)] \right\} e^{-E^2/4\sigma_x^2} \\ \langle V^2(t) \rangle &= 2\sigma_x^2 + E^2 \end{aligned}\right\} \qquad (11.35)$$

Consequently, the variance is

$$\sigma_{out}^2 = \langle V^2 \rangle - V_0^2 = 2\sigma_x^2 + E^2 - V_0^2 \qquad (11.36)$$

Since when there is no signal (i.e., when $E = 0$) the d-c component, in accordance with (11.35) and (11.18), is equal to $\sqrt{\pi/2}\sigma_x$, the increment of the d-c component, caused by the signal, will be $V_0 - \sqrt{\pi/2}\sigma_x$. Consequently, the power signal-to-noise ratio at the linear detector output is

$$(S/N)_{out} = (V_0 - \sqrt{\pi/2}\,\sigma_x)^2 / (2\sigma_x^2 + E^2 - V_0^2) \qquad (11.37)$$

Let us consider the extreme cases where $(E^2/2\sigma_x^2) \ll 1$ (low-level signal) and $(E^2/2\sigma_x^2) \gg 1$ (high-level signal). For a low-level signal, omitting the terms with powers of $(E^2/2\sigma_x^2)$ higher than the first [which are obtained in asymptotic representation of the functions $I_0(E^2/4\sigma_x^2)$ and $I_1(E^2/4\sigma_x^2)$], expression (11.35) can be written in short as follows:

$$V_0 \approx \sqrt{\pi/2}\,\sigma_x \left(1 + \frac{1}{2} E^2/2\sigma_x^2\right)$$

In this case, the increment of the d-c component is

$$V_0 - \sqrt{\pi/2}\,\sigma_x = \frac{1}{2}\sqrt{\pi/2}\,\sigma_x E^2/2\sigma_x^2$$

and the variance

$$\sigma_{out}^2 = 2\sigma_x^2 + E^2 - V_0^2 \approx 2\sigma_x^2 + E^2 - \frac{\pi}{2}\sigma_x^2 \left(1 + \frac{E^2}{2\sigma_x^2}\right)$$

$$= \left(1 - \frac{\pi}{4}\right) 2\sigma_x^2 \left(1 + \frac{E^2}{2\sigma_x^2}\right) \approx \left(1 - \frac{\pi}{4}\right) 2\sigma_x^2$$

Thus, for $(E^2/2\sigma_x^2) \ll 1$, the signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_{out} = \frac{\frac{\pi}{8}\sigma_x^2\left(\frac{E^2}{2\sigma_x^2}\right)^2}{\left(1-\frac{\pi}{4}\right)2\sigma_x^2} \approx \left(\frac{E^2}{2\sigma_x^2}\right)^2 \approx \left(\frac{S}{N}\right)_{in}^2 \qquad (11.38)$$

For a high-level signal, the approximate expressions of Bessel functions of large arguments permit (11.35) to be represented in the form

$$V_0 \approx \sigma_x \sqrt{2}\frac{E}{\sqrt{2}\sigma_x}\left[1+\frac{1}{4(E^2/2\sigma_x^2)} + \frac{1}{32}\frac{1}{(E^2/2\sigma_x^2)^2}+\ldots\right] \approx E\left(1+\frac{\sigma_x^2}{2E^2}\right) \qquad (11.39)$$

Substituting this value of V_0 into formula (11.37), we get for $(E^2/2\sigma_x^2) \gg 1$

$$\left(\frac{S}{N}\right)_{out} \approx \frac{E^2}{\sigma_x^2} \approx 2\frac{E^2}{2\sigma_x^2} \approx 2\left(\frac{S}{N}\right)_{in} \qquad (11.40)$$

Let us consider square-law detection in a similar way.

Substituting $V(t)$ for $A(t)$ in formula (11.25), we obtain the output voltage of the square-law detector

$$v_{out}(t) = K\left[E^2/2 + A^2(t)/2 + EA(t)\cos\theta(t)\right] \qquad (11.41)$$

Averaging this expression over time and taking into account that $\overline{A^2(t)} = 2\sigma_x^2$ and $\overline{A(t)\cos\theta(t)} = 0$ {as well as the average $\overline{x(t)} = \overline{A(t)\cos[\omega_0 t + \theta(t)]}$}, we obtain the d-c component of the output voltage of the square-law detector:

$$\overline{v_{out}(t)} = K\left(\frac{E^2}{2}+\sigma_x^2\right) = V_{0s}+V_{0n} \qquad (11.42)$$

The term $V_{0n} = K\sigma_x^2$ defines the d-c component produced by noise [see expression (11.27)] when there is no signal. On the other hand, the term $V_{0s} = KE^2/2$, which is the increment of the d-c component due to the effect of the harmonic signal voltage, can be regarded as the useful signal at the detector output.

Squaring expression (11.41), we obtain

$$v_{out}^2(t) = K^2\left[\frac{E^2}{2}+\frac{A^2(t)}{2}+EA(t)\cos\theta(t)\right]^2$$

$$= K^2\left[\frac{E^4}{4}+\frac{A^4(t)}{4}+E^2A^2(t)\left(\frac{1}{2}+\frac{1}{2}\cos 2\theta(t)\right)\right.$$

$$\left.+\frac{E^2A^2(t)}{2}+E^3A(t)\cos\theta(t)+A^3(t)E\cos\theta(t)\right]$$

On averaging, the terms containing $\cos\theta(t)$ and $\cos 2\theta(t)$ vanish. Therefore, the average output power* is

$$\overline{v_{out}^2(t)} = K^2\left[\frac{E^4}{4} + \frac{1}{4}\overline{A^4(t)} + E^2\overline{A^2(t)}\right]$$

$$= K^2\left(\frac{E^4}{4} + 2\sigma_x^4 + 2E^2\sigma_x^2\right) \qquad (11.43)$$

Subtracting $\overline{(v_{out})}^2$ from this expression, we find the variance of noise at the output of the square-law detector:

$$\sigma_{out}^2 = K^2\left(\frac{E^4}{4} + 2\sigma_x^4 + 2E^2\sigma_x^2\right)$$

$$- K^2\left(\frac{E^4}{4} + E^2\sigma_x^2 + \sigma_x^4\right) = K^2(E^2\sigma_x^2 + \sigma_x^4) \qquad (11.44)$$

For $E = 0$, this expression turns into (11.29). Now let us set up the expression for the power signal-to-noise ratio at the detector output:

$$\left(\frac{S}{N}\right)_{out} = \frac{V_{0s}^2}{\sigma_{out}^2} = \frac{K^2\left(\frac{E^4}{4}\right)}{K^2(E^2\sigma_x^2 + \sigma_x^4)} = \frac{\left(\frac{E^2}{2\sigma_x^2}\right)^2}{1 + 2\frac{E^2}{2\sigma_x^2}} \qquad (11.45)$$

But $E^2/2\sigma_x^2$ is the power signal-to-noise ratio at the detector input. Thus, for $(S/N)_{in} \ll 1$ (i.e., for $E^2/2 \ll \sigma_x^2$)

$$\left(\frac{S}{N}\right)_{out} \approx \left(\frac{S}{N}\right)_{in}^2 \qquad (11.46)$$

while for high values of $(S/N)_{in}$, i.e., for $E^2/2 \gg \sigma_x^2$,

$$\left(\frac{S}{N}\right)_{out} \approx \frac{1}{2}\left(\frac{S}{N}\right)_{in} \qquad (11.47)$$

For example, for $E^2/2\sigma_x^2 = 1/10$, the ratio $(S/N)_{out} = 1/120$ [formula (11.45)], while for $E^2/2\sigma_x^2 \geqslant 4$, the ratio $(S/N)_{out}$ is close to a half of $(S/N)_{in}$.

Formula (11.45) allows one to draw the following important conclusion: in the case of a *low-level* (with respect to noise) *signal*, in the square-law detector there occurs *suppression of the signal*, while in that of a *high-level signal*, the ratio $(S/N)_{out}$ is *proportional* to $(S/N)_{in}$.

* On averaging $A^4(t)$, we get

$$\langle A^4(t)\rangle = \int_0^\infty A^4 p(A)\,dA = 8\sigma_x^4$$

Since the process in question is ergodic, in this section no difference is made between time and ensemble averaging.

Let us compare the results obtained for square-law and linear detection. Comparison of formulas (11.46) and (11.38) shows that in the case of a low-level signal and strong noise the linear and square-law detectors behave in a similar way: the signal-to-noise ratio at the output is proportional to the square of the signal-to-noise ratio at the input. Thus, in the *linear detector a low-level signal is also suppressed.* Analysis shows that this property is inherent in detectors with any other current-voltage characteristics.

On the other hand, for $E \gg \sigma_x$, the power signal-to-noise ratio at the output of the square-law detector is only one fourth of that at the output of the linear detector [compare formulas (11.47) and (11.40)]. This is explained by the fact that in square-law detection a high-level signal carries the noise to a characteristic curve section with a greater slope, this causing a relative increase in noise. Indeed, let us assume that the envelope of an input harmonic oscillation, equal to 1 V, has got an increment $a \ll 1$. Then the output voltage of the square-law detector will increase, according to (11.25), from $K/2$ to $(K/2)(1 + a)^2 \approx (K/2)(1 + 2a)$, i.e., the relative increment will be $2a$, while in linear detection this increment will be only a. Transforming from voltage to power, we obtain the above-mentioned four-fold loss.

Although the above analysis refers to a harmonic (unmodulated) signal, the obtained conclusions can be extended to cover the processing of rectangular radio pulses in the midst of random noise, when a pulse at the detector output is an increment of the d-c component of the rectified voltage over a time interval equal to the pulse duration.

The amplitude modulation of the signal, that can be regarded as a *slow* variation of the d-c component of the detector output voltage, also has no significant influence on the comparative evaluation of square-law and linear detection.

Finally, it should be noted that all the results obtained in this section do not depend on the relationship between the signal carrier frequency ω_0 and the instantaneous noise frequency $\omega_0 + \dot{\theta}$.

From this it follows that the superposition of frequency or phase modulation on the signal (with a constant amplitude) does not affect the signal-to-noise ratio at the detector output. This is in agreement with the basic properties of the amplitude detector, revealed in Ch. 8.

11.6. JOINT EFFECT OF A HARMONIC OSCILLATION AND NORMAL NOISE ON A FREQUENCY DETECTOR

On the basis of the principle of operation of a frequency detector, discussed in Sec. 8.9, in the text below we shall proceed from the block diagram shown in Fig. 11.10. The signal $s(t)$ at the input of the amplitude limiter is a frequency-modulated wave (we mean tone

modulation)

$$s(t) = A_s \cos\left(\omega_0 t + \frac{\omega_d}{\Omega} \sin \Omega t\right) \tag{11.48}$$

while noise is a random normal process with an energy spectrum $W_x(\omega) = W_0$ that is uniform within the passband of the I.F. filter (in a superheterodyne receiver).

The passband $2\Delta\omega_0$ of this filter can be taken to be equal to twice the frequency deviation, i.e., $\Delta\omega_0 = \omega_d$. The low-pass filter at the

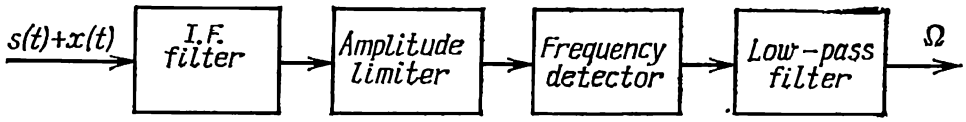


Fig. 11.10. Block diagram of a frequency detector

detector output must have a passband from 0 to Ω_{max}, where Ω_{max} is the highest modulation frequency. Noise acting at the limiter input will be written, as in the preceding section, in the form $x(t) = A(t) \cos[\omega_0 t + \theta(t)]$.

When analyzing the joint effect of $s(t)$ and $x(t)$ on the frequency detector, we can ease the problem by considering separately the following two operating conditions: (1) operation in the absence of the useful frequency modulation, in which case at the detector input there exist a purely harmonic oscillation $s(t) = A_s \cos \omega_0 t$ and noise $x(t)$, and (2) operation in the presence of frequency modulation. We shall assume that in the second case noise at the detector output remains the same as in the first case.

Thus, in the absence of modulation, the combined oscillation at the limiter input is [see (11.31)]

$$s(t) + x(t) = A_s \cos \omega_0 t + A(t) \cos[\omega_0 t + \theta(t)] = V(t) \cos[\omega_0 t + \xi(t)] \tag{11.49}$$

where $V(t)$ and $\xi(t)$ are defined by expressions (11.32) and (11.33).

Denoting the limitation threshold by V_{thr}, we come to the following expression for the oscillation at the limiter output:

$$v_{out}(t) = V_{thr} \cos[\omega_0 t + \xi(t)] \tag{11.50}$$

[compare with (8.45)].

The voltage at the output of the frequency detector, which is proportional to the derivative of the phase $\xi(t)$ in the absence of the useful modulation, is noise. Thus,

$$x_{out}(t) = S_{fd} \dot{\xi}(t) \tag{11.51}$$

where S_{fd} is the slope of the characteristic curve of the frequency detector (see Sec. 8.9). One can see that the intensity and structure of noise $x_{out}(t)$ at the output of the frequency detector is completely defined by the statistical characteristics of the derivative of the phase $\dot{\xi}(t)$.

The general expression for the phase for any relations between $A(t)$ and $A_s(t)$ has the form of (11.33). However, under actual conditions of reception of frequency-modulated oscillations, the level of the signal is considerably higher than that of noise. As a rule, $A_s^2/2\sigma_x^2 \gg 1$. (As in the preceding section, σ_x^2 is the average power of noise at the detector input). Therefore, expression (11.33) for the phase can be simplified:

$$\xi(t) \approx \arctan\left[\frac{A(t)\sin\theta(t)}{A_s}\right] \approx \frac{A(t)}{A_s}\sin\theta(t) \qquad (11.52)$$

The statistical characteristics of the random function $\xi(t) = [A(t)/A_s]\sin\theta(t)$ coincide with the characteristics found in Sec. 4.6 for the quadrature components of a narrow-band process. It was shown there that the function $A(t)\sin\theta(t)$ possesses a normal distribution and an energy spectrum $2W_x(\omega_0 + \Omega)$ [see expression (4.64)]. Thus,

$$W_\xi(\Omega) = 2W_x(\omega_0 + \Omega)/A_s^2 \qquad (11.53)$$

When differentiating a normal random process, its distribution remains normal (see Sec. 7.1). Consequently, $\dot{\xi}(t)$, i.e., the instantaneous frequency deviation, is also normally distributed.

Thus, for $E^2/2\sigma_x^2 \gg 1$, *noise at the output of the frequency detector (as at the input) is a normal random process.*

Now we must determine the energy spectrum of the process $\dot{\xi}(t)$. For this purpose, it is sufficient to multiply $W_\xi(\Omega)$ by Ω^2 (see Sec. 7.3). Thus,

$$W_{\dot{\xi}}(\Omega) = \Omega^2 W_\xi(\Omega) = \frac{2\Omega^2}{A_s^2} W_x(\omega_0 + \Omega) \qquad (11.54)$$

while the energy spectrum of noise at the output of the frequency detector, in accordance with (11.51), is

$$W_{out}(\Omega) = S_{fd}^2 W_{\dot{\xi}}(\Omega) = \frac{2S_{fd}^2\Omega^2}{A_s^2} W_x(\omega_0 + \Omega) \qquad (11.55)$$

Finally, the autocorrelation function of noise at the output of the low-pass filter (with a passband Ω_{max}) is

$$B_{out}(\tau) = \frac{1}{2\pi}\int_{-\Omega_{max}}^{\Omega_{max}} W_{out}(\Omega)\, e^{i\Omega\tau}\, d\Omega$$

$$= \frac{2S_{fd}^2}{2\pi A_s^2}\int_{-\Omega_{max}}^{\Omega_{max}} \Omega^2 W_x(\omega_0 + \Omega)\, e^{i\Omega\tau}\, d\Omega \qquad (11.56)$$

and the variance, i.e., the average power, of noise

$$\sigma_{out}^2 = B_{out}(0) = \frac{S_{fd}^2}{\pi A_s^2} \int_{-\Omega_{max}}^{\Omega_{max}} \Omega^2 W_x(\omega_0 + \Omega)\, d\Omega \qquad (11.57)$$

Now let us consider the conditions of frequency modulation, under which the voltage at the detector output is proportional to the frequency deviation. In tone modulation, the voltage amplitude is

$$V_s = S_{fd}\omega_d \qquad (11.58)$$

Thus, the signal power at the output (without taking into account the effect of noise) is equal to $V_s^2/2 = S_{fd}^2\omega_d^2/2$, while the power of noise (neglecting modulation) is defined by expression (11.57). Consequently, the signal-to-noise ratio at the output

$$\left(\frac{S}{N}\right)_{out}^{fm} = \frac{V_s^2/2}{\sigma_{out}^2} = \frac{\omega_d^2 A_s^2}{\frac{2}{\pi}\int_{-\Omega_{max}}^{\Omega_{max}} \Omega^2 W_x(\omega_0 + \Omega)\, d\Omega} \qquad (11.59)$$

Let us illustrate expression (11.59) by the following example. Let the interfering signal at the detector input be white noise with an energy spectrum $W_x(\omega) = W_0 = \text{const}$.

Then the integral in (11.59) is equal to $2\Omega_{max}^3 W_0/3$ and expression (11.59) is easily reduced to the form

$$\left(\frac{S}{N}\right)_{out}^{fm} = \frac{(A_s^2/2)\, 3\omega_d^2}{(1/\pi)\, 2\Omega_{max}^3 W_0} = 3\,\frac{A_s^2/2}{W_0 2\,(2F_{max})}\left(\frac{\omega_d}{\Omega_{max}}\right)^2$$

But $A_s^2/2$ is the signal power at the input, while $W_0 2\,(2F_{max})$ is none other than σ_x^2, i.e., the power of noise in two bands $2\Delta f_0 = 2F_{max}$ (one in the region $\omega > 0$ and the other in the region $\omega < 0$).

Thus, finally

$$\left(\frac{S}{N}\right)_{out}^{fm} = 3\left(\frac{\omega_d}{\Omega_{max}}\right)^2 \left(\frac{S}{N}\right)_{in} \qquad (11.60)$$

By increasing the ratio ω_d/Ω_{max}, i.e., the angle modulation index, one can obtain a large gain in the signal-to-noise ratio as compared with the signal-to-noise ratio in systems with amplitude modulation. This process is widely used in VHF broadcasting systems and also in the sound channels of telecasting systems.

It should be emphasized that the advantages of wide-band frequency modulation obtain so long as the signal is stronger than noise and the amplitude of the oscillation at the detector input is limited. If noise is stronger than the signal, the latter is suppressed.

11.7. INTERACTION OF A HARMONIC OSCILLATION AND NORMAL NOISE IN AN AMPLITUDE LIMITER WITH A RESONANT LOAD

As in Sec. 11.5, let the sum of a harmonic oscillation $s(t) = E\cos\omega_0 t$ and narrow-band noise $x(t) = A(t)\cos[\omega_0 t + \theta(t)]$ be represented in the form

$$v(t) = V(t)\cos[\omega_0 t + \xi(t)]$$

where the slowly varying functions $V(t)$ (envelope) and $\xi(t)$ (phase) are determined by expressions (11.32) and (11.33).

The current-voltage characteristic of the nonlinear element effecting the amplitude limitating is defined by the expression

$$i(v) = \begin{cases} av^{\nu} & \text{for } v > 0 \\ 0 & \text{for } v \leqslant 0 \end{cases} \qquad (11.61)$$

where a is a constant coefficient, ν is a positive real number, and i and v are normalized (dimensionless) quantities.

The curves of $i(v)$ for $\nu = 0$, 1/2, 1, and 2 are shown in Fig. 11.11. The current i is nonzero only for positive half-waves of the input oscillation. The cases $\nu = 1$ and 2 correspond to half-wave rectification (detection); $\nu = 1$ also corresponds to the optimum operating conditions of a nonlinear resonance amplifier with a cutoff angle $\theta = 90°$ (see Sec. 8.4). The case $\nu = 1/2$ is typical of the same amplifier operating with v within the saturation region of the characteristic curve, and also of soft amplitude limiting. Finally, the case $\nu = 0$ corresponds to ideal amplitude limiting.

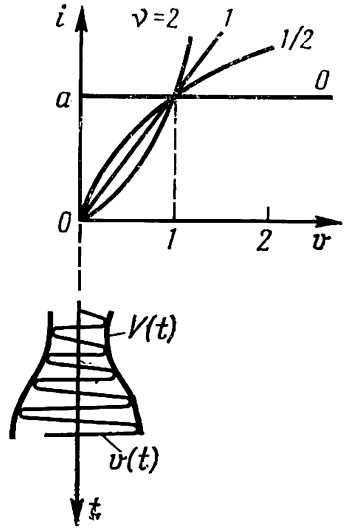


Fig. 11.11. Characteristics of nonlinear elements

The general expression for the current is

$$i(t) = \sum_{m=0}^{\infty} C(\nu, m)\, V^{\nu}(t) \times \cos[m\omega_0 t + m\xi(t)] \qquad (11.62)$$

This expression determines the current $i(t)$ as a sum of oscillations whose centre frequencies $m\omega_0$ and phases $m\xi(t)$ are multiples of the frequency ω_0 and phase $\xi(t)$. The envelopes $V^{\nu}(t)$ of these oscillations are equal to the νth power of the excitation amplitude. The constant coefficients $C(\nu, m)$ depend on the exponent ν, i.e., on the shape of the current-voltage characteristic $i(v)$.

If the system is provided with an output low-pass filter, only the term corresponding to $m = 0$ should be taken into account:

$$i_{l-f}(t) = C(\nu, 0)\, V^{\nu}(t)$$

For a linear detector, i.e., for $\nu = 1$, we have

$$i_{l-f}(t) = C(1, 0)\, V(t)$$

and for a square-law detector (half-wave detector), i.e., for $\nu = 2$

$$i_{l-f}(t) = C(2, 0)\, V^2(t)$$

As would be expected, in the first case the output quantity is proportional to the first power and in the second, to the square of the envelope of the input voltage.

In the case of a limiter with a selective load that separates a frequency band in the vicinity of ω_0, for the term corresponding to $m = 1$, we obtain

$$i_{\omega_0}(t) = C(\nu, 1)\, V^{\nu}(t) \cos[\omega_0 t + \xi(t)] \qquad (10.63)$$

From this expression stems an important conclusion about the structure of the oscillation at the output of a resonance amplifier: *the frequency and phase of the oscillation fully coincide with the frequency and phase of the input signal.* The law governing the change of the envelope $V^{\nu}(t)$ depends on the shape of the current-voltage characteristic.

For $\nu = 0$ (ideal limiter), the amplitude of the input oscillation is constant, so that expression (11.63) assumes the form

$$i_{\omega_0}(t) = C(0, 1) \cos[\omega_0 t + \xi(t)] \qquad (11.63')$$

The coefficients $C(\nu, m)$ in expression (11.63) are defined by representing the characteristic $i(v)$ in the form of a contour integral*. Let us give the values of the three coefficients entering into the previous expressions:

$$C(1, 0) = a/\pi,\ C(2, 0) = a/4,\ C(0, 1) = 2a/\pi$$

These coefficients can also be obtained by analyzing a harmonic excitation of a nonlinear element (and a narrow-band noise within one period is close to harmonic).

In fact, under the conditions of linear half-wave detection, illustrated in Fig. 11.11, i.e., with a current cutoff angle of 90°, the mean value of the current is equal to $1/\pi$ of the pulse amplitude. With a quadratic characteristic of the detector, it is equal to 1/4. The amplitude of the first harmonic in the case of a complete amplitude limiting, where the current pulses have a rectangular shape, is equal to $2/\pi$ of the pulse amplitude.

The spectral composition of the current $i(t)$ in the case $\nu = 0$ (ideal limiting) is shown in Fig. 11.12. Figure 11.12*a* shows the ener-

* See[3].

gy spectrum $W_x(\omega)$ of noise $x(t)$ at the limiter input and the spectral power density $2\pi\frac{E^2}{2}\delta(\omega \pm \omega_0)$ of the harmonic oscillation.

Figure 11.12*b* shows the spectral bands of noise and the discrete spectral lines of the harmonic components of the current $i(t)$ (accurate to within the scale factors). In the case of a narrow-band input noise and a harmonic signal, whose spectra are concentrated near the

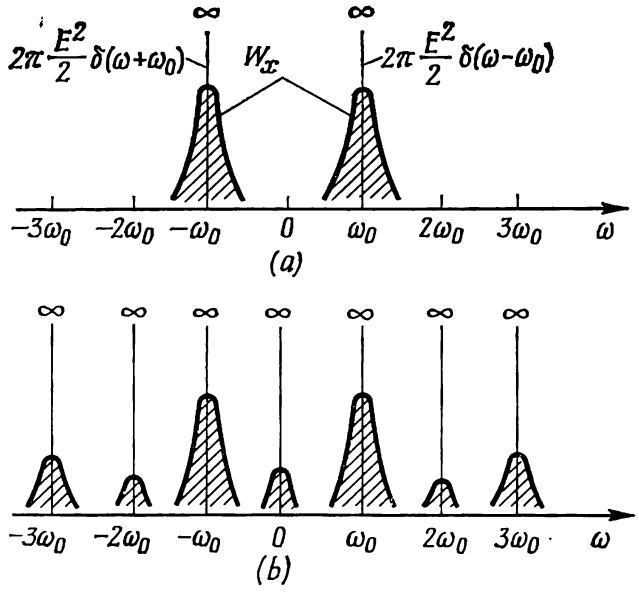


Fig. 11.12. Spectrum at (*a*) the input and (*b*) the output of an ideal limiter

frequency ω_0, the spectrum of the current $i(t)$ is concentrated near the frequencies $\omega = 0, \pm\omega_0, \pm 2\omega_0, \pm 3\omega_0$, etc.

Since the limiter load is of resonant character, the energy spectrum $W_{out}(\omega)$ must be taken into account only in the bands that are concentrated near the frequencies $\pm\omega_0$ (Fig. 11.12). Consequently, the limiter current producing a voltage across the resonant load can be represented in the form of the sum

$$y(t) = A_0 \cos \omega_0 t + i_n(t)$$

where $i_n(t)$ is the narrow-band noise oscillation with a centre frequency ω_0.

A detailed analysis [3] shows that the power of the fluctuation component $i_n(t)$ in the main spectral band (near $\omega \pm \omega_0$) amounts to about 80% of the total power of the output noise.

The relation between the power of the useful oscillation, equal to $\frac{1}{2}A_0^2$, and the power of noise $i_n(t)$ is of primary interest.

A characteristic feature of an ideal limiter is the independence of the total power (average) of the combined oscillation $y(t)$ of the power signal-to-noise ratio at the limiter input. In any case this power is $P_y = \frac{1}{2}(2a/\pi)^2$, where a is the constant (see Fig. 11.11) determining the amplitude of the rectangular diode current pulses.

When there is no useful signal ($E = 0$), the entire power P_y is concentrated in noise. When there is no noise at the input, the entire power P_y is concentrated in the signal, in which case the amplitude A_0 reaches its maximum value $A_{0,\max} = 2a/\pi$.

When $s(t)$ and $x(t)$ act simultaneously, the amplitude A_0 is defined by the expression

$$A_0 = \frac{a}{\sqrt{2\pi}\sigma_x} E e^{-E^2/4\sigma_x^2}\left[I_0\left(\frac{E^2}{4\sigma_x^2}\right)+I_1\left(\frac{E^2}{4\sigma_x^2}\right)\right]$$

$$= \frac{a}{\sqrt{\pi}} h e^{-h^2/2}\left[I_0\left(\frac{h^2}{2}\right)+I_1\left(\frac{h^2}{2}\right)\right] \qquad (11.64)$$

where I_0 and I_1 are Bessel functions of imaginary argument (see Sec. 11.5) and $h^2 = E^2/2\sigma_x^2$ is the signal-to-noise ratio at the limiter input.

It should be noted that in the absence of noise, when $h \to \infty$ {see [6], formula (8.451.5)}

$$I_0\left(\frac{h^2}{2}\right) = I_1\left(\frac{h^2}{2}\right) \to \frac{e^{h^2/2}}{\sqrt{2\pi}\,(h/\sqrt{2})} = \frac{e^{h^2/2}}{\sqrt{\pi}\,h}$$

and expression (11.64) gives $A_{0,\max} = 2a/\pi$.

In the other extreme case $h^2 \ll 1$ (low-level signal), where

$$I_0(h^2/2) \to 1,\quad I_1(h^2/2) \to h^2/4,\quad e^{-h^2/2} \to 1$$

the signal amplitude

$$A_0 \approx (a/\sqrt{\pi})h$$

The ratio

$$\frac{A_0}{A_{0,\max}} = \frac{\sqrt{\pi}}{2} h e^{-h^2/2}\left[I_0\left(\frac{h^2}{2}\right)+I_1\left(\frac{h^2}{2}\right)\right]$$

can be regarded as the coefficient of attenuation of the signal by noise in the limiter.

Let us give the power signal-to-noise ratio.

In the case of a low-level signal [3] ($h^2 \ll 1$)

$$\left(\frac{S}{N}\right)_{out} = \frac{A_0^2}{A_{0,\max}^2} = \frac{\left(\frac{ah}{\sqrt{\pi}}\right)^2}{\left(\frac{2a}{\pi}\right)^2} = \frac{\pi}{4}h^2 = \frac{\pi}{4}\left(\frac{S}{N}\right)_{in}$$

while in that of a high-level signal ($h^2 \gg 1$)

$$\left(\frac{S}{N}\right)_{out} = 2\left(\frac{S}{N}\right)_{in}$$

Thus, for $h^2 \ll 1$, the relative attenuation of the signal in the limiter is equal to only $\pi/4$. With a high-level signal ($h^2 \gg 1$), the output signal-to-noise ratio is twice the input one. It should be noted that the relative decrease in the variance of noise at the output is due to the suppression of the component which is in phase with the signal. The variance of the orthogonal noise component causing the fluctuation of the phase of the output signal is not reduced.

11.8. AUTOCORRELATION FUNCTION AND ENERGY SPECTRUM OF A RANDOM PROCESS IN A PARAMETRIC CIRCUIT

Let the transfer function of a linear parametric circuit be a real function of time, independent of frequency. In Sec. 10.2 it was shown that such a transfer function characterizes a circuit in which amplitude modulation occurs.

Let us designate the transfer function by $K(t)$ (the argument $i\omega$ is omitted), the function $K(t)$ being such that it may be either a deterministic or a random process. The input signal $s(t)$ may also be either a deterministic or a random process.

Let us set up an expression for the autocorrelation function of the output signal $s_{out}(t)$:

$$\begin{aligned} B_{out}(t, \tau) &= \langle s_{out}(t)\, s_{out}(t+\tau)\rangle \\ &= \langle K(t)\, K(t+\tau)\, s(t)\, s(t+\tau)\rangle \end{aligned} \tag{11.65}$$

We are interested in the case where the transfer function $K(t)$ is independent of the input signal $s(t)$. Then the average value of the product in (11.65) is equal to the product of the average values of the respective factors, i.e.,

$$\begin{aligned} B_{out}(t, \tau) &= \langle K(t)\, K(t+\tau)\rangle \langle s(t)\, s(t+\tau)\rangle \\ &= B_K(t, \tau)\, B_s(t, \tau) \end{aligned} \tag{11.66}$$

where $B_s(t, \tau)$ is the autocorrelation function of the input signal, while

$$B_K(t, \tau) = \langle K(t)\, K(t+\tau)\rangle \tag{11.67}$$

is the *autocorrelation function of the circuit* with a transfer factor $K(t)$.

From expression (11.66) follows an important property of a linear circuit with variable parameters: *the autocorrelation function of the output signal is equal to the product of the autocorrelation function of the input signal*, $B_s(t,\tau)$, *and that of the circuit*, $B_K(t,\tau)$.

In the case of nonstationary processes, the autocorrelation functions in (11.66) and (11.67) depend not only on the time shift τ, but

also on the time t. These characteristics are not always convenient to use. In the examples given later in the text we shall use the functions $B(\tau)$ obtained by averaging $B(t, \tau)$ over t [see Sec. 4.7, formula (4.89)].

Applying the Fourier transformation to the time-averaged function $B_{out}(\tau)$, we obtain the energy spectrum (also averaged) of the output signal:

$$W_{out}(\omega) = \int_{-\infty}^{\infty} B_{out}(\tau)\, e^{-i\omega\tau}\, d\tau \tag{11.68}$$

Let us illustrate relations (11.65) through (11.68) by examples.

1. A harmonic signal $s(t) = \cos \omega_0 t$ acts at the input of a linear circuit with a transfer function

$$K(t) = K_0 + \Delta K(t) \tag{11.69}$$

where K_0 is the average value of the circuit gain and $\Delta K(t)$ is the gain fluctuation which is a normally distributed stationary random process with a variance σ_K^2.

To obtain a complete characteristic of the time variation of the transfer function of the circuit, either the autocorrelation function $B_K(\tau)$ or the energy spectrum $W_K(\omega)$ of the random process $K(t)$ must be specified.

It is clear that to the d-c component K_0 corresponds the energy spectrum $W_{K0}(\omega) = 2\pi K_0^2 \delta(\omega)$*. Let us specify the energy spectrum of the second term, i.e., of $\Delta K(t)$, in the form

$$W_{\Delta K}(\omega) = 2c/(a^2 + \omega^2)$$

where a and c are constants.

Thus, the energy spectrum of the sum $K_0 + \Delta K(t)$ is

$$W_K(\omega) = 2\pi K^2 \delta(\omega) + 2c/(a^2 + \omega^2) \tag{11.70}$$

To the given energy spectrum $W_K^0(\omega)$ corresponds the autocorrelation function

$$B_K(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_K(\omega)\, e^{i\omega\tau}\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} 2\pi K_0^2 \delta(\omega)\, e^{i\omega\tau} d\omega$$

$$+ \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2c}{a^2+\omega^2}\, e^{i\omega\tau} d\omega = K_0^2 + \frac{c}{a}\, e^{-a|\tau|} \tag{11.71}$$

Let us find the autocorrelation function and the energy spectrum of the signal at the circuit output.

Having in view relations (11.66) and (11.71) and also taking into account that the autocorrelation function of the signal $s(t) = \cos \omega_0 t$ is

$$B_s(\tau) = \frac{1}{2} \cos \omega_0 \tau$$

* In fact, for the d-c component K_0, the autocorrelation function is equal to K_0^2. Consequently, according to formula (11.68), the energy spectrum

$$W_{K0}(\omega) = K_0^2 \int_{-\infty}^{\infty} e^{-i\omega\tau}\, d\tau$$

$$= 2\pi K_0^2 \delta(\omega)$$

we obtain

$$B_{out}(\tau)=B_K(\tau)\,B_s(\tau)=\frac{1}{2}\left(K_0^2+\frac{c}{a}\,e^{-a|\tau|}\right)\cos\omega_0\tau \qquad (11.72)$$

Now, let us find the energy spectrum by means of expression (11.68):

$$W_{out}(\omega)=\frac{1}{2}\int_{-\infty}^{\infty}\left(K_0^2+\frac{c}{a}\,e^{-a|\tau|}\right)\cos\omega_0\tau e^{-i\omega\tau}d\tau$$

$$=\frac{1}{4}K_0^2\left[\int_{-\infty}^{\infty}e^{-i(\omega-\omega_0)\tau}d\tau+\int_{-\infty}^{\infty}e^{-i(\omega+\omega_0)\tau}d\tau\right]$$

$$+\frac{1}{4}\frac{c}{a}\left[\int_{-\infty}^{\infty}e^{-a|\tau|}e^{-i(\omega-\omega_0)\tau}d\tau+\int_{-\infty}^{\infty}e^{-a|\tau|}e^{-i(\omega+\omega_0)\tau}d\tau\right]$$

The first two integrals give the delta functions $2\pi\delta(\omega-\omega_0)$ and $2\pi\delta(\omega+\omega_0)$. The last two integrals give respectively $2a/[a^2+(\omega-\omega_0)^2]$ and $2a/[a^2+(\omega+\omega_0)^2]$. Thus, finally, we have

$$W_{out}(\omega)=\frac{\pi}{2}K_0^2\left[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)\right]+\frac{c}{2}\left[\frac{1}{a^2+(\omega-\omega_0)^2}+\frac{1}{a^2+(\omega+\omega_0)^2}\right] \qquad (11.73)$$

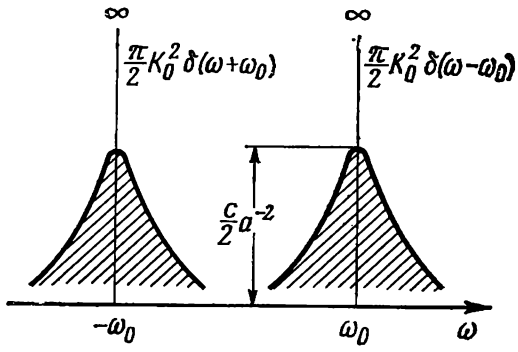


Fig. 11.13. Energy spectrum at the output of a parametric circuit with a random transfer function in the case of a harmonic excitation

The function $W_{out}(\omega)$ is shown in Fig. 11.13. The two discrete spectral lines correspond to the monochromatic component of the output signal, while the continuous spectrum (shaded in Fig. 11.13) corresponds to the noise component caused by the gain fluctuation $\Delta K(t)$. This spectrum consists of combination frequencies arranged symmetrically with respect to the signal frequency ω_0 (in the region of negative ω, symmetrically with respect to $-\omega_0$).

2. A normally distributed random process $s(t)$, whose energy spectrum (Fig. 11.14)

$$W_s(\omega)=2d/(b^2+\omega^2) \qquad (11.74)$$

is concentrated near the zero frequency, acts at the input of a circuit with a transfer function

$$K(t)=K_0(1+M\cos\Omega t)\ \text{for}\ M<1 \qquad (11.75)$$

Let us find the autocorrelation function of the input signal

$$B_s(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{2d}{b^2+\omega^2}\,e^{i\omega\tau}\,d\omega=\frac{d}{b}e^{-b|\tau|} \qquad (11.76)$$

and the autocorrelation function of the circuit

$$B_K(\tau) = K_0^2 + \frac{1}{2} M^2 K_0^2 \cos \Omega\tau \tag{11.77}$$

Then, in accordance with relation (11.66), the autocorrelation function of the output signal is

$$B_{out}(\tau) = B_K(\tau)\, B_s(\tau) = \frac{d}{b}\left(K_0^2 + \frac{1}{2} M^2 K_0^2 \cos \Omega\tau \right) e^{-b|\tau|} \tag{11.78}$$

and the energy spectrum

$$W_{out}(\omega) = \frac{d}{b} K_0^2 \int_{-\infty}^{\infty} e^{-b|\tau|} e^{-i\omega\tau}\, d\tau + \frac{dM^2K_0^2}{2b} \int_{-\infty}^{\infty} e^{-b|\tau|} \cos \Omega\tau e^{-i\omega\tau}\, d\tau$$

$$= K_0^2 \frac{2d}{b^2+\omega^2} + K_0^2 M^2 d \left[\frac{1}{b^2+(\omega-\Omega)^2} + \frac{1}{b^2+(\omega+\Omega)^2} \right] \tag{11.79}$$

The function $W_{out}(\omega)$ is shown in Fig. 11.14.

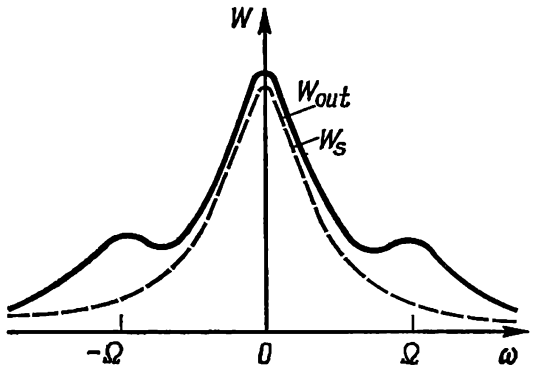


Fig. 11.14. Energy spectrum at the output of a parametric circuit with a harmonically varying transfer function, which is acted upon by a normal random process

3. A normally distributed random process $s(t)$ acts at the input of a circuit whose transfer function $K(t)$ is also a random process with normal distribution.

Let us give the energy spectra of the processes $s(t)$ and $K(t)$ in the form $W_s(\omega) = 2d/(b^2 + \omega^2)$, as in example 2, and $W_{K_i}(\omega) = 2\pi K_0^2 \delta(\omega) + 2c/(a^2 + \omega^2)$, as in example 1.

The autocorrelation functions of the input signal and the circuit under consideration are respectively

$$B_s(\tau) = (d/b)\, e^{-b|\tau|}, \quad B_K(\tau) = K_0^2 + (c/a)\, e^{-a|\tau|}$$

Now, we find the autocorrelation function of the output signal to be

$$B_{out}(\tau) = B_K(\tau)\, B_s(\tau) = \frac{d}{b} K_0^2 e^{-b|\tau|} + \frac{cd}{ab} e^{-(a+b)|\tau|} \tag{11.80}$$

and the energy spectrum

$$W_{out}(\omega) = K_0^2 \frac{2d}{b^2+\omega^2} + \frac{cd}{ab} \frac{2(a+b)}{[(a+b)^2+\omega^2]} \tag{11.81}$$

The first term on the right-hand side corresponds to the signal at the output of the circuit with a transfer function K_0 (in the absence of multiplicative noise)

and the second term corresponds to multiplicative noise. The magnitude of this term is proportional to the product of the parameter d characterizing the signal intensity and the parameter c that determines the variance σ_K^2 of the fluctuation of the transfer function.

11.9. EFFECT OF MULTIPLICATIVE NOISE ON THE DISTRIBUTION LAW OF A SIGNAL

The correlation and spectral characteristics of a random signal, considered in the preceding section, are not exhaustive. The determination of the probability density $p(s_{out})$ is of great practical interest.

In the general case, where the transfer function $K(i\omega, t)$ of a circuit is a function of two variables—frequency and time — it is very difficult to find $p(s_{out})$ if the distribution law of the input signal is arbitrary. The problem is much simplified in the case of multiplicative noise of amplitude modulation type, where the transfer function $K(t)$ depends only on a single variable, namely, the time t.

Keeping this in mind, let us consider the following three typical situations:

(1) $s(t)$ is a random process and $K(t)$ is a deterministic process;
(2) $s(t)$ is a deterministic process and $K(t)$ is a random process;
(3) both $s(t)$ and $K(t)$ are random processes.

Situations 1 and 2 lead to the problem of finding the distribution law of the product of $s(t)$ and $K(t)$, in which one of the factors is a random and the other, a deterministic variable. If the random process is stationary, the problem can easily be solved. It is known from the theory of random functions that multiplying a random function $x(t)$ (stationary process) with a differential distribution $p(x)$, zero mean, and variance σ_x^2 by a deterministic time function $y(t)$ yields a nonstationary process $x(t)\,y(t)$ with the same distribution law but with variance $\sigma_{out}^2 = \sigma_x^2 y^2(t)$.

In particular, if the input signal $s(t)$ is a stationary normally distributed process with a variance σ_s^2, whereas the transfer function $K(t)$ of the system is deterministic (case 1), the output signal retains normal distribution, but to each fixed instant of time corresponds its own dispersion $\sigma_{out}^2 = \sigma_s^2 K^2(t)$.

In the case of a deterministic signal $s(t)$ and a random function $K(t)$ (case 2), if the latter can be represented in the form $K(t) = K_0 + \Delta K(t)$, it is advisable to write the output signal in the form

$$s_{out}(t) = K(t)\,s(t) = K_0 s(t) + \Delta K(t)\,\dot{s}(t) = s_{out.det}(t) + s_{out.ran}(t) \tag{11.82}$$

The first term on the right-hand side characterizes the useful output signal (deterministic), while the second, multiplicative noise

(random). The distribution law of this term is the same as that of the random process $\Delta K(t)$, but with a variance $\sigma_K^2 s^2(t)$ (for $\overline{\Delta K(t)} = 0$).

Let us consider case 3. Let both the processes $s(t)$ and $K(t)$ be stationary random processes with probability densities $p(s)$ and $p(K)$, respectively. The problem is to find the probability density of the random process $s_{out}(t)$ which is the product of $s(t)$ and $K(t)$.

From the theory of probability it is known that if two mutually independent random variables x and y have probability densities $p(x)$ and $p(y)$, their product $z = xy$ will have the probability density $p(z)$ defined by the expression

$$p(z) = \int_{-\infty}^{\infty} p(x)\, p\left(\frac{z}{x}\right) \frac{dx}{|x|} \tag{11.83}$$

Implying by x the input signal $s(t)$, by y the transfer function $K(t)$ and by z the product $s_{out}(t) = s(t)\, K(t)$, we obtain an expression for determining the probability density of the output signal $s_{out}(t)$.

Let us illustrate the application of (11.83) by the following examples.

1. As a first example, let us consider the transmission of a harmonic signal $s(t) = A_0 \cos(\omega_0 t + \theta)$, whose epoch angle θ is a random variable uniformly distributed over an interval $(-\pi, \pi)$, through a linear circuit with a transfer function $K(t)$ that fluctuates with respect to its mean value K_0 according to a normal law.

Thus, the respective probability densities are

$$p(s) = \frac{1}{\pi \sqrt{A_0^2 - s^2}}, \qquad -A_0 < s < A_0$$

$$p(K) = \frac{1}{\sqrt{2\pi}\,\sigma_K} e^{-(K-K_0)^2/2\sigma_K^2}, \qquad -\infty < K < \infty$$

Substituting these expressions into (11.83) and setting $x = s$ and $z/x = y = s_{out}/s = K$, we come to the following general expression for the probability density of the output signal

$$p(s_{out}) = \frac{1}{\pi \sqrt{2\pi}\,\sigma_K} \int_{-A_0}^{A_0} \frac{\exp\left[-\dfrac{(s_{out}/s - K_0)^2}{2\sigma_K^2}\right]}{\sqrt{A_0^2 - s^2}} \frac{ds}{|s|}$$

It should be emphasized that the distribution law thus found characterizes the *instantaneous value* of the output signal.

In practice, the distribution of the envelope of the output signal is often of principal interest.

Representing the output signal in the form

$$S_{out}(t) = s(t)\, K(t) = A_0 \cos(\omega_0 t + \theta)\,[K_0 + \Delta K(t)] = A(t) \cos(\omega_0 t + \theta)$$

where $A(t) = A_0 [K_0 + \Delta K(t)]$ is the envelope, we arrive at the obvious conclusion that the randomness of the phase angle θ has no effect on the distribution of the envelope, which coincides with the distribution of the function $K(t)$, i.e., it is a normal distribution with a mean $A_0 K_0$ and a variance $A_0^2 \sigma_K^2$.

2. Now, let us analyse the statistical characteristics of a *normally distributed* random signal $s(t)$ transmitted through a circuit whose transfer function $K(t)$ is also a normally distributed random variable.

To simplify the analysis, let us represent the transfer function in the form $K(t) = K_0 + \Delta K(t)$, so that the output signal can be written in the form of a sum similar to (11.82):

$$s_{out}(t) = K_0 s(t) + \Delta K(t)\, s(t) = s'_{out}(t) + s_{mn}(t)$$

The first term on the right-hand side, that differs from $s(t)$ by the constant coefficient K_0, characterizes the useful signal at the circuit output. The statistical properties of this signal coincide with those of the random signal $s(t)$ at the circuit input, and its variance is equal to $K_0^2\sigma_s^2$.

Let us consider the term $s_{mn}(t)$ which is a result of the action of multiplicative noise. By hypothesis,

$$p(s) = \frac{1}{\sqrt{2\pi}\,\sigma_s} e^{-s^2/2\sigma_s^2} \quad \text{for} \quad -\infty < s < \infty$$

$$p(\Delta K) = \frac{1}{\sqrt{2\pi}\,\sigma_K} e^{-\Delta K^2/2\sigma_K^2} \quad \text{for} \quad -\infty < \Delta K < \infty$$

Substituting these expressions into (11.83), we find the probability density of noise $s_{mn}(t)$:

$$p(s_{mn}) = \frac{1}{2\pi\sigma_s\sigma_K} \int_{-\infty}^{\infty} e^{-s^2/2\sigma_s^2} e^{-\frac{s_{mn}^2/s^2}{2\sigma_K^2}} \frac{ds}{|s|}$$

$$= \frac{1}{\pi\sigma_s\sigma_K} \int_0^{\infty} \exp\left\{-\left[\frac{s^2}{2\sigma_s^2} + \frac{s_{mn}^2}{2\sigma_K^2 s^2}\right]\right\} \frac{ds}{s} \qquad (11.84)$$

The integral in this expression is a particular case of the integral {see [6], formula (3.478.4)}

$$\int_0^{\infty} x^{\nu-1} \exp(-\beta x^p - \gamma x^{-p})\, dx = \frac{2}{p} (\gamma/\beta)^{\nu/2p} K_{\nu/p}(2\sqrt{\beta\gamma}), \qquad (11.85)$$

$$(\operatorname{Re}\beta > 0, \operatorname{Re}\gamma > 0)$$

Here $K_{\nu/p}(z)$ are cylindrical functions of imaginary argument.

Setting [in accordance with (11.84)] $\nu = 0$, $\beta = 1/2\sigma_s^2$, $\gamma = s_{mn}^2/2\sigma_K^2$, and $p = 2$, we obtain

$$p(s_{mn}) = \frac{1}{\pi\sigma_K\sigma_s} K_0\left(2\sqrt{\frac{1}{2\sigma_s^2}\frac{s_{mn}^2}{2\sigma_K^2}}\right) = \frac{1}{\pi\sigma_s\sigma_K} K_0\left(\frac{|s_{mn}|}{\sigma_s\sigma_K}\right) \qquad (11.86)$$

The designation K_0 of the cylindrical function of order zero should not be confused with that of the mean of the transfer function $K(t)$ [see (11.69)].

To calculate the function $K_0(z)$, one can use the series {see [6], formula (8.447.3)}

$$K_0(z) = -\ln\frac{z}{2} J_0(z) + \sum_{k=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{2k}}{(k!)^2} \psi(k+1) \qquad (11.87)$$

where $J_0(z)$ is a Bessel function of imaginary argument and ψ is Euler's function

$$\psi(k+1) = -C + \sum_{n=1}^{k} \frac{1}{n}$$

where $C = 0.5772$ is the Euler constant.

Taking into account (11.87), expression (11.86) transforms into the following one:

$$p(s_{mn}) = \frac{1}{\pi\sigma_K\sigma_s}\left[-\ln\frac{|s_{mn}|}{2\sigma_K\sigma_s} I_0\left(\frac{|s_{mn}|}{\sigma_s\sigma_K}\right) + \sum_{k=0}^{\infty}\left(\frac{|s_{mn}|}{2\sigma_s\sigma_K}\right)^{2k}\frac{1}{(k!)^2}\psi(k+1)\right] \quad (11.88)$$

The graph of the function $\sigma_s\sigma_K(s_{mn})$ is shown in Fig. 11.15. This graph is of generalized character that illustrates the distribution law of the product of

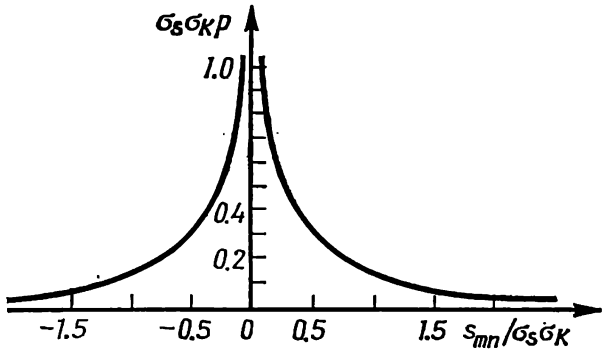


Fig. 11.15. Probability density of a product of two independent normal random processes

two normal, mutually independent random processes with variances σ_s^2 and σ_K^2 and zero means.

3. Assume that the input signal is a narrow-band normally distributed process

$$s(t) = A(t)\cos[\omega_0 t + \theta(t)], \quad (11.89)$$

the envelope $A(t)$ and the function $\Delta K(t)$ having energy spectra of approximately the same width.

In this case, it is convenient to represent the oscillation $s_{mn}(t)$ in a form similar to (11.89):

$$s_{mn}(t) = A(t)\,\Delta K(t)\cos[\omega_0 t + \theta(t)] = A_{out}(t)\cos[\omega_0 t + \theta(t)]$$

In practice, it is often necessary to know the distribution law of the envelope $A_{out}(t)$, i.e., of the random function

$$A_{out}(t) = A(t)\Delta K(t)$$

Taking into account that A has a Rayleigh distribution

$$p(A) = \frac{A}{\sigma_s^2} e^{-A_s^2/2\sigma_s^2}, \quad 0 < A < \infty,$$

8163927

while $\Delta K(t)$ has a normal distribution

$$p(\Delta K)=p\left(\frac{A_{out}}{A}\right)=\frac{1}{\sqrt{2\pi}\,\sigma_K}e^{-A_{out}^2/2\sigma_K^2A^2},\quad -\infty<\Delta K<\infty,$$

and using (11.83), we obtain

$$p(A_{out})=\int_0^\infty p(A)\,p\left(\frac{A_{out}}{A}\right)\frac{dA}{A}=\frac{1}{\sqrt{2\pi}\,\sigma_s^2\sigma_K}\int_0^\infty e^{\frac{-A^2}{2\sigma_s^2}}e^{\frac{-A_{out}^2}{2\sigma_K^2}\frac{1}{A^2}}dA$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_s^2\sigma_K}\int_0^\infty e^{-\beta A^2-\gamma/A^2}\,dA, \quad (11.90)$$

$$\beta=1/2\sigma_s^2,\quad \gamma=A_{out}^2/2\sigma_K^2$$

The integral in expression (11.90) is a particular case of the integral (11.85)

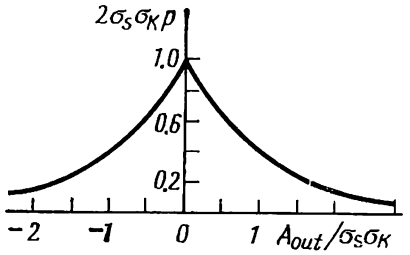


Fig. 11.16. Probability density of the envelope of a random process at the output of a parametric circuit with a random transfer function, which is acted upon by a normal process

for $\gamma = 1$ and $p = 2$.

Thus,

$$p(A_{out})=\frac{1}{\sqrt{2\pi}\,\sigma_s^2\sigma_K}\left(\frac{\gamma}{\beta}\right)^{1/4}K_{1/2}(2\sqrt{\beta\gamma})$$

Taking into account that {see [6], formula (8.469.3)}

$$K_{1/2}(z)=\sqrt{\frac{\pi}{2z}}e^{-z}$$

and substituting β and γ, we obtain

$$p(A_{out})=\frac{1}{2\sigma_s\sigma_K}e^{\frac{-|A_{out}|}{\sigma_s\sigma_K}} \quad (11.91)$$

The graph of the function $2\sigma_s\sigma_K p(A_{out})$ is shown in Fig. 11.16. In contrast to the envelope $A(t)$ that cannot be negative, quantity $A_{out}(t)$, being the product of $A(t)$ and $\Delta K(t)$, has an exponential distribution symmetrical with respect to zero.

Chapter 12

MATCHED FILTERING OF SIGNALS ON A NOISY BACKGROUND

12.1. GENERAL

When synthesizing radio circuits for the detection and processing of signals on a noisy background, first of all, one has to find an optimum transfer function of the circuit. The approach to this problem depends on the designation of the given device. For example, the following problems may have to be considered.

1. Detection of a signal, when it is necessary to find out whether the received oscillation contains the useful signal or only noise.
2. Evaluation of parameters where it is necessary to determine to the highest possible accuracy (as to the mean-square error) one or several parameters of the useful signal, such as amplitude, frequency, etc. (obviously, this can be done only after this signal has been detected so that it is reliably observed at the receiver output).
3. Discrimination of signals where several signals can exist at the input and it is required to indicate which of these signals are present, some apriori information being available on each of these signals.
4. Reproduction of the original shape of a signal that has been distorted due to the effect of noise.
5. Prediction (extrapolation) of a signal where it is necessary to forecast its most probable future values on the basis of its "history".

The problems of detection, evaluation of parameters, and discrimination arise, for example, in radar engineering, radio astronomy, and radio measurements. The problems of reproduction and prediction are typical of automatic control.

The methods of finding algorithms describing such conversions of the excitation as ensure optimum solution of the above problems are described in special courses. The study of possibilities of attenuation of the harmful effect of noise where the signal *is known* is of primary importance in the theory of radio circuits and signals. Therefore, the next sections (12.2 through 12.4) deal with the determination of parameters of filters which are optimal for the given signal contaminated with noise. Such filters are referred to as matched to the signal. Some examples of constructing such filters are given in Sec. 12.5.

12.2. MATCHED FILTERING OF A GIVEN SIGNAL

The linear filtering of a signal in order to separate it from a mixture "signal plus noise" is one of the fundamental processes effected in any radio receiver. The filtering is based on the utilization of the frequency selectivity of oscillatory circuits. During the first 50 to 60 years of the development of radio engineering frequency filters of this type were required to transmit as uniformly as possible the signal spectrum and to suppress as fully as possible the frequencies lying outside this spectrum. A filter with a rectangular, sharp-cutoff amplitude-frequency characteristic was considered ideal.

With the development of the theory of information and the statistical theory of signal detection both the determination of the functions of the linear filter and the approach to its construction have changed.

It became obvious that the former approach had two shortcomings: (1) it did not take into account the shape of the signal (that may be different with the same width of the signal spectrum) and (2) it disregarded the statistical properties of noise. Therefore, even the ideal filter with the sharp-cutoff amplitude-frequency characteristic failed to provide for the maximum signal-to-noise ratio at the output, since it was not matched to the signal and noise.

The works of N. Wiener, A. Kolmogorov, V. Kotelnikov and others marked a major breakthrough in the theory and practice of linear filtering. For the first time these scientists were able to solve the problem of synthesizing an optimal filter for a given signal contaminated with noise of specified statistical characteristics.

The optimization criteria may be different, depending on which of the above-given problems has to be solved. For the problem of the detection of a signal in noise, the criterion of the maximum signal-to-noise ratio at the filter output has gained the widest application. The filters meeting this criterion are called *matched.* The present chapter deals exclusively with such filters.

The requirements for a filter maximizing the signal-to-noise ratio may be expressed as follows. An additive mixture of a signal $s(t)$ and noise $n(t)$ is applied to the input of a linear two-port network with constant parameters and a transfer function $\mathbf{K}(i\omega)$ (Fig. 12.1). The signal is completely known; this means that its shape and position on the time axis are specified. Noise is a random process with specified statistical characteristics. It is necessary to synthesize a filter providing for the maximum possible ratio of the peak value of the signal to the root-mean-square value of noise at the output, no conditions being imposed as to the preservation of the true shape of the signal, since the shape is not important for the detection of the signal in noise.

To understand the essence of the matched filtering of a signal, let us first consider a simplest case where only the useful signal $s(t)$ with the known spectrum $\mathbf{S}(\omega) = S(\omega)e^{i\varphi_s(\omega)}$ acts at the input of a filter with a uniform amplitude-frequency characteristic. It is necessary to find the phase-frequency characteristic of the filter, maximizing the peak of the signal at the filter output. It should be

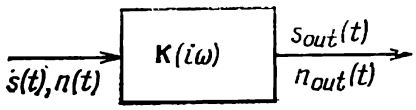


Fig. 12.1. Effect of a signal and noise on a linear two-port network

recalled that the spectral density $\mathbf{S}(\omega)$ of the signal completely defines its energy E. No change of the phase relations in the spectrum has any effect on the signal energy (see Sec. 2.8). Therefore, the above problem is equivalent to the problem of maximization of the peak of the output signal at the *given energy of the input signal.*

Let us represent the output signal in the form

$$s_{out}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{S}(\omega)\,\mathbf{K}(i\omega)\,e^{i\omega t}\,d\omega \tag{12.1}$$

where

$$\mathbf{K}(i\omega) = K_0 e^{i\varphi_K(\omega)} \tag{12.2}$$

is the transfer function of the nonminimal-phase two-port network with the sought for phase-frequency characteristic $\varphi_K(\omega)$ and the uniform amplitude-frequency characteristic $K_0 = \text{const}$.

Thus,

$$s_{out}(t) = \frac{K_0}{2\pi}\int_{-\infty}^{\infty} \mathbf{S}(\omega)\,e^{i[\omega t+\varphi_K(\omega)]}\,d\omega$$

$$= \frac{K_0}{2\pi}\int_{-\infty}^{\infty} S(\omega)\,e^{i[\omega t+\varphi_s(\omega)+\varphi_K(\omega)]}\,d\omega \tag{12.3}$$

Using the obvious inequality

$$\int_a^b F(x)\,dx \leqslant \int_a^b |F(x)|\,dx \tag{12.4}$$

and taking into account that $|e^{i[\omega t+\varphi_s(\omega)+\varphi_K(\omega)]}| = 1$, we can set up the following inequality:

$$s_{out}(t) \leqslant \frac{K_0}{2\pi}\int_{-\infty}^{\infty} |\mathbf{S}(\omega)|\,d\omega \tag{12.5}$$

This inequality defines the upper limit of the instantaneous value of the oscillation $s_{out}(t)$ for the given spectrum of the input signal. The maximum peak of the output voltage is obtained when inequality (12.5) turns into equality, for which purpose it is necessary, as it follows from comparison between expressions (12.3) and (12.5), to ensure a certain definite relation between the phase characteristic $\varphi_K(\omega)$ of the filter and the phase characteristic $\varphi_s(\omega)$ of the spectrum of the input signal.

Assume that the output signal reaches its maximum at a moment t_0 (as yet indefinite). In this case, expression (12.3) yields

$$s_{out,\max}(t) = s_{out}(t_0) = \frac{K_0}{2\pi}\int_{-\infty}^{\infty} S(\omega)\, e^{i[\omega t_0+\varphi_s(\omega)+\varphi_K(\omega)]}\, d\omega \quad (12.6)$$

and the condition for inequality (12.5) to transform into equality is reduced to the following:

$$\omega t_0 + \varphi_s(\omega) + \varphi_K(\omega) = 0$$

or

$$\varphi_K(\omega) = -[\varphi_s(\omega) + \omega t_0] \quad (12.7)$$

This relation may be called the *condition for compensation of the epoch angles* in the signal spectrum. Indeed, the first term on the right-hand side of (12.7), equal to $-\varphi_s(\omega)$, compensates for the phase characteristic $\varphi_s(\omega)$ of the input spectrum $S(\omega)$. When the signal passes through a filter with a phase characteristic $\varphi_K(\omega)$, all the phase-corrected components of the spectrum are added together to form the output signal peak at the instant of time $t = t_0$.

The relations between the phase characteristic $\varphi_s(\omega)$ of the input spectrum, the phase-compensating characteristic $-\varphi_s(\omega)$, and the complete phase characteristic of the filter $\varphi_K(\omega) = -[\varphi_s(\omega) + \omega_0 t_0]$ are illustrated in Fig. 12.2.

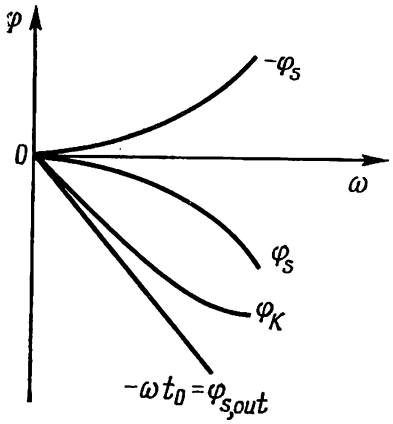


Fig. 12.2. Relations between the phase characteristics of a signal spectrum at the input and output of a matched filter

Having passed through the filter, the output signal spectrum will have the phase characteristic

$$\varphi_{s,out}(\omega) = \varphi_s(\omega) + \varphi_K(\omega) = \varphi_s(\omega) + [-\varphi_s(\omega) - \omega t_0] = -\omega t_0 \quad (12.8)$$

shown in the same figure.

The physical meaning and the minimum possible value of t_0 are considered in greater detail in the next section, however, from simple physical concepts it is clear that for the peak to be formed, the

entire energy of the signal must be utilized, and this is possible not before the action of the input signal has ceased. In other words, the delay t_0 cannot be shorter than the total duration of the signal.

Now, let us extend the analysis to include noise at the filter input. With the energy spectrum $W(\omega)$ of noise being uniform, i.e., $W(\omega) = W_0 = \text{const}$ (white noise), a filter with a uniform amplitude-frequency characteristic is completely unacceptable, because the noise power at the output reaches an extremely high (theoretically infinitely high) value.

To find the optimum transfer function *maximizing the signal-to-noise ratio* at the filter output, let us set up expressions for the signal and noise, first separately, and then in the form of their ratio.

The peak value of the signal is determined by the expression

$$s_{out,\,\max}(t) = s_{out}(t_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{S}(\omega)\,\mathbf{K}(i\omega)\,e^{i\omega t_0}\,d\omega \qquad (12.9)$$

The root-mean-square value of noise at the filter output is given by

$$\sigma = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} W(\omega)\,K^2(\omega)\,d\omega} = \sqrt{\frac{W_0}{2\pi}\int_{-\infty}^{\infty} K^2(\omega)\,d\omega}$$

Consequently, the ratio of the peak value of the signal to the root-mean-square value of noise at the filter output is

$$\frac{s_{out}(t_0)}{\sigma} = \frac{\left|\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{S}(\omega)\,\mathbf{K}(i\omega)\,e^{i\omega t_0}\,d\omega\right|}{\left(\frac{W_0}{2\pi}\right)^{1/2}\left[\int_{-\infty}^{\infty} K^2(\omega)\,d\omega\right]^{1/2}} \qquad (12.10)$$

Let us use the well-known Schwarz inequality

$$\left|\int_a^b \mathbf{F}_1(x)\,\mathbf{F}_2(x)\,dx\right|^2 \leqslant \int_a^b |\mathbf{F}_1(x)|^2\,dx \int_a^b |\mathbf{F}_2(x)|^2\,dx \qquad (12.11)$$

where $\mathbf{F}_1(x)$ and $\mathbf{F}_2(x)$ are generally complex functions.

This inequality turns into equality only if the following condition is satisfied:

$$F_2(x) = A\,\mathbf{F}_1^*(x) \qquad (12.12)$$

i.e., if $\mathbf{F}_2(x)$ is proportional to the complex conjugate of $\mathbf{F}_1(x)$ (A is an arbitrary constant coefficient).

Using the symbols of (12.9) and setting $\mathbf{S}(\omega) = \mathbf{F}_1(i\omega)$ and $\mathbf{K}(i\omega)e^{i\omega t_0} = \mathbf{F}_2(i\omega)$, we may rewrite expression (12.11) in the

form

$$\left|\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}(\omega)\,\mathbf{K}(i\omega)\,e^{i\omega t_0}\,d\omega\right|^2$$

$$\leqslant\frac{1}{2\pi}\int_{-\infty}^{\infty}S^2(\omega)\,d\omega\,\frac{1}{2\pi}\int_{-\infty}^{\infty}K^2(\omega)\,d\omega \qquad (12.13)$$

Hence, the quantity $s_{out}(t_0)/\sigma$ determined by expression (12.10) satisfies the condition

$$\frac{s_{out}(t_0)}{\sigma}\leqslant\frac{\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}S^2(\omega)\,d\omega\,\frac{1}{2\pi}\int_{-\infty}^{\infty}K^2(\omega)\,d\omega\right]^{1/2}}{\left[\frac{W_0}{2\pi}\int_{-\infty}^{\infty}K^2(\omega)\,d\omega\right]^{1/2}}=$$

$$=\frac{1}{\sqrt{W_0}}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}S^2(\omega)\,d\omega\right]^{1/2} \qquad (12.14)$$

Taking into account that the expression in the brackets is none other than the total energy E of the input signal [see (2.66)], let us write the last expression in the form

$$s_{out}(t_0)/\sigma\leqslant\sqrt{E/W_0} \qquad (12.15)$$

Finally, from expression (12.12) it follows that this inequality turns into equality only if

$$\mathbf{K}(i\omega)\,e^{i\omega t_0}=A\mathbf{S}^*(\omega)$$

whence

$$\mathbf{K}(i\omega)=A\mathbf{S}^*(\omega)\,e^{-i\omega t_0} \qquad (12.16)$$

The obtained relation completely defines the transfer function maximizing the signal-to-noise ratio at the filter output.

Taking into account that $\mathbf{S}(\omega)=S(\omega)e^{i\varphi_S(\omega)}$ and the complex conjugate $\mathbf{S}^*(\omega)=S(\omega)e^{-i\varphi_S(\omega)}$, let us rewrite expression (12.16) in the following form:

$$\mathbf{K}(i\omega)=K(\omega)\,e^{i\varphi_K(\omega)}=AS(\omega)\,e^{-i[\varphi_S(\omega)+\omega t_0]} \qquad (12.17)$$

From this relation stem the following two conditions for the phase and amplitude characteristics of a matched filter:

$$\varphi_K(\omega)=-[\varphi_s(\omega)+\omega t_0] \qquad (12.18)$$

$$K(\omega)=AS(\omega) \qquad (12.19)$$

Where by the complex transfer function is implied a dimensionless quantity (e.g., the ratio between the complex voltage amplitudes

at the input and output), the dimension of the coefficient A must be the inverse of that of the spectral density of the signal.

Relations (12.18) and (12.19) have a profound physical meaning. The first of them, coinciding with (12.7) and defining the *compensation of the epoch angles in the signal spectrum*, was discussed above. Relation (12.19) establishing the fact that the spectral density modulus $K(\omega)$ must have the same form as the spectral density modulus

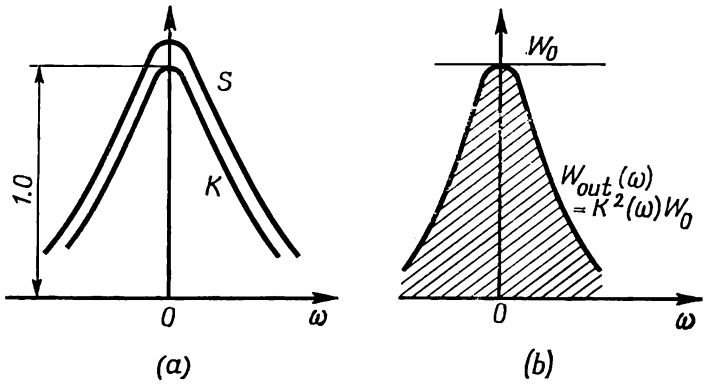


Fig. 12.3. (*a*) Spectral density of a signal and the amplitude-frequency characteristic of a matched filter and (*b*) the energy spectra at the input and output

$S(\omega)$ of the signal can also be easily interpreted from the standpoint of physics. With the amplitude-frequency characteristic $K(\omega)$ satisfying condition (12.19), the filter lets pass through the spectral components of noise nonuniformly, attenuation being inversely proportional to the modulus $S(\omega)$. This considerably reduces the noise power at the filter output. In Fig. 12.3*b* this power is determined by the area (shaded) under the curve $W_{out}(\omega) = K^2(\omega)W_0$. (To make the presentation more graphic, the characteristic curves in Fig. 12.3 are plotted under assumption that $AS(0) = 1$.)

The attenuation of the signal due to the nonuniformity of the characteristic $K(\omega)$ is less pronounced than the attenuation of noise, because the decrease of $K(\omega)$ affects the spectral components whose contribution to the signal peak is comparatively small. As a result, there occurs the *attenuation of noise relative to the signal.* In conjunction with the phase compensation of the signal spectrum (this compensation having no effect on the variance of the output noise), this results in that the *signal-to-noise ratio at the filter output is maximized.*

12.3. IMPULSE RESPONSE OF THE MATCHED FILTER. PHYSICAL REALIZABILITY

The fact that the transfer factor $\mathbf{K}(i\omega)$ of the matched filter is the complex conjugate of the signal spectrum $\mathbf{S}(\omega)$ also points to a close relation between the time characteristics of the filter and

signal. To reveal this relation, let us find the impulse response of the matched filter.

Using expression (5.28) and taking into account formula (12.16), we obtain

$$g(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{K}(i\omega)\,e^{i\omega t}\,d\omega=A\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}^*(\omega)\,e^{i\omega(t-t_0)}\,d\omega \qquad (12.20)$$

Taking into account that

$$\mathbf{S}^*(\omega)=\mathbf{S}(-\omega)$$

and transforming to a new variable $\omega_1 = -\omega$, we rewrite expression (12.20) as follows:

$$g(t)=-\frac{A}{2\pi}\int_{+\infty}^{-\infty}\mathbf{S}(\omega_1)\,e^{-i\omega_1(t-t_0)}\,d\omega_1$$

$$=A\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{S}(\omega_1)\,e^{i\omega_1(t_0-t)}\,d\omega_1 \qquad (12.21)$$

The right-hand side of this expression is none other than the function $As(t_0 - t)$. Consequently, if the signal $s(t)$ is specified, the impulse response $g(t)$ of the matched (optimum) filter is defined as the function

$$g(t)=As(t_0-t) \qquad (12.22)$$

The graph of the function $s(t_0 - t)$ is plotted in Fig. 12.4. The curve $s(-t)$ is a mirror image of the given signal $s(t)$, with the

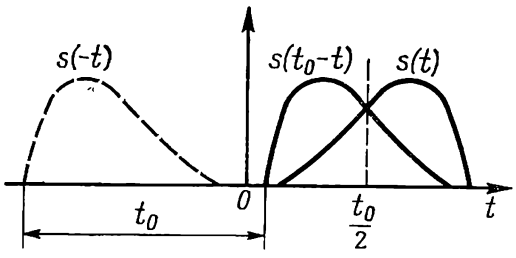


Fig. 12.4. Plotting the function which is a mirror image of a given signal

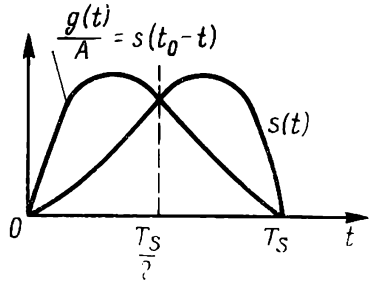


Fig. 12.5. Plotting the impulse response curve of a matched filter

symmetry axis coinciding with the ordinate axis. The function $s(t_0 - t)$ shifted to the right by an amount t_0 with respect to $s(-t)$ is also a mirror image of the original signal $s(t)$, but with the symmetry axis passing through the point $t_0/2$ on the abscissa axis. Shown in Fig. 12.5 is a similar plot for the case where the time is read from the beginning of the signal.

Since the impulse response of a physical circuit cannot start at $t < 0$ [the filter response cannot lead the stimulus $\sigma(t)$], it is clear that the delay t_0 in expression (12.16) cannot be shorter than T_s. It is only for $t_0 \geqslant T_s$ that the energy of the signal can be used to the full to produce the maximum possible peak at the point $t = t_0$. It is clear that an increase of t_0 in excess of T_s has no effect on the peak of the output signal, but simply shifts it to the right (towards delay).

On the other hand, the condition $t_0 \geqslant T_s$ requires that the *duration T_s of the signal $s(t)$ be finite*; only in this case it is possible to realize the peak of the signal, with the delay t_0 being finite. In other words, the use of matched filtering to maximize the signal-to-noise ratio in the above-discussed sense is possible for a *pulse* signal (and a time-limited pulse train as well).

Now, let us study the question of the physical realizability of the matched filter. Let us have an arbitrary signal $s(t)$ to which correspond the impulse response $g(t)$ of a matched filter and the Fourier transform $\mathbf{K}(i\omega)$ of this function determined by expressions (12.22) and (12.16), respectively. A question arises: under what conditions $\mathbf{K}(i\omega)$ can be the transfer function of a *physically realizable two-port network*.

The answer to this question is given by the Paley-Wiener realizability criterion, according to which the inequality

$$\int_0^\infty \frac{|\ln K(\omega)|}{1+\omega^2}\, d\omega < \infty \tag{12.23}$$

is the necessary condition for the positive function $K(\omega)$ to be the modulus of the transfer function of a passive electric circuit. Since in the problem under consideration $K(\omega) = AS(\omega)$ [see (12.19)], condition (12.23) can be written in the form*

$$\int_0^\infty \frac{|\ln S(\omega)|}{1+\omega^2}\, d\omega < \infty \tag{12.24}$$

i.e., integral (12.24) must be convergent.

It should be pointed out that condition (12.24) is necessary but not sufficient. In fact, condition (12.24) is reduced to the requirement that the duration of the signal $s(t)$ should be limited. For example, in the case of the signal spectrum determined by expression (2.82) [see Fig. 2.20*b*] the time function $s(t)$ is not limited in time and so is the required impulse response $g(t) = As(t_0 - t)$ of the filter, that cannot be realized. In this case, integral (12.24) diverges, because outside the interval $(-\omega_m, \omega_m)$ the function $S(\omega)$ vanishes while $|\ln S(\omega)|$ goes to infinity.

* Here by ω we imply a dimensionless quantity.

Although the Paley-Wiener criterion leaves open the question of the circuit structure, it yields some useful data about the properties of realizable circuits.

In particular, from this criterion it follows that the amplitude-frequency characteristic $K(\omega)$ must be quadratically integrable, i.e., $\int_{-\infty}^{\infty} K^2(\omega)\, d\omega < \infty$. Further, the amplitude-frequency characteristic $K(\omega)$ can be zero only for some discrete frequencies, but not within a finite frequency range. In fact, if within the range $\omega_1 < \omega < \omega_2$ the function $K(\omega) = 0$, $|\ln K(\omega)|$ goes to infinity and, as mentioned above, the integral diverges, i.e., condition (12.23) is not satisfied. From this it follows that filters with a sharp-cutoff amplitude-frequency characteristic cannot be realized, although in practice it is possible to obtain characteristics very close to ideal.

It is also obvious that the transfer function $K(\omega) = Ke^{-\alpha\omega}$, $\omega > 0$, is realizable, since $|\ln K(\omega)| = |\ln K - \alpha\omega|$ rises slower than the denominator in (12.23), i.e., condition (12.23) is satisfied. On the other hand, a Gaussian filter with a transfer function $K(\omega) = Ke^{-\alpha\omega^2}$ cannot be realized, since $|\ln K(\omega)| = |\ln K - \alpha\omega^2|$ increases with increasing ω at the same rate as the denominator in (12.23).

12.4. SIGNAL AND NOISE AT THE OUTPUT OF THE MATCHED FILTER

In order to find the shape of the output signal, let us use the general expression

$$s_{out}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, \mathbf{K}(i\omega)\, e^{i\omega t}\, d\omega$$

Substituting into this expression relation (12.16), we obtain

$$s_{out}(t) = A\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}(\omega)\, \mathbf{S}^*(\omega)\, e^{-i\omega t_0} e^{i\omega t}\, d\omega$$

$$= A\frac{1}{2\pi} \int_{-\infty}^{\infty} S^2(\omega)\, e^{i\omega(t-t_0)}\, d\omega \qquad (12.25)$$

For $t = t_0$, this expression transforms into

$$s_{out}(t_0) = AE \qquad (12.26)$$

where E is the energy of the input signal.

On the other hand, from the first part of relation (12.25), which contains the product of the spectral densities $\mathbf{S}(\omega)$ and

$S^*(\omega)e^{-i\omega t_0}$, it follows that $s_{out}(t)$ can be represented as the convolution of the following two functions of time: $s(t)$ and $s(t_0-t)$.

Thus,

$$s_{out}(t) = A\int_{-\infty}^{\infty} s(t-x)\, s(t_0-x)\, dx$$

Substituting $t - x = y$ and $t_0 - x = y - (t - t_0)$, we have

$$s_{out}(t) = A\int_{-\infty}^{\infty} s(y)\, s[y-(t-t_0)]\, dy \qquad (12.27)$$

One can see that the integral on the right-hand side of this expression is none other than the *autocorrelation function* $B_s(t - t_0)$ of the input signal. Thus, we come to an important conclusion that

$$s_{out}(t) = AB_s(t - t_0) \qquad (12.28)$$

Thus, the signal at the output of the matched filter *coincides with the autocorrelation function of the input signal* to within the constant coefficient A.

To plot the graph of the function $s_{out}(t)$, given the function $B_s(\tau)$, it is sufficient to replace τ in this function by $t - t_0$ (and take into account the coefficient A). For $t = t_0$, i.e., for $\tau = 0$, the value $B_s(0)$ is equal to the signal energy. Therefore, as in the spectral analysis, the peak value of the signal [see (12.26)] is

$$s_{out}(t_0) = AB_s(0) = AE$$

Now let us consider the parameters and statistical characteristics of noise at the output of the matched filter. In the case of white noise with a normal distribution (it is exactly this type of noise that is of principal practical interest), the noise distribution at the output of a linear filter remains normal. The energy spectrum of noise at the filter output, as is clear from (7.1) and Fig. 12.3, is $W_{out}(\omega) = K^2(\omega) W_0$. Consequently, the autocorrelation function of noise at the output of the matched filter is

$$B_{out}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W_{out}(\omega)\, e^{i\omega\tau}\, d\omega = \frac{W_0}{2\pi}\int_{-\infty}^{\infty} K^2(\omega)\, e^{i\omega\tau}\, d\omega$$

Substituting $K(\omega) = AS(\omega)$ and taking into account expressions (12.25) and (12.28), in which we set $t - t_0 = \tau$, we obtain

$$B_{out}(\tau) = A^2W_0\frac{1}{2\pi}\int_{-\infty}^{\infty} S^2(\omega)\, e^{i\omega\tau}\, d\omega$$

$$= AW_0 s_{out}(t_0+\tau) = AW_0AB_s(\tau) = A^2W_0B_s(\tau) \qquad (12.29)$$

From this it follows that the autocorrelation function of noise at the output of the matched filter coincides in form with the autocor-

relation function of the input signal (and, therefore, with the output signal).

Setting $\tau = 0$, we find the variance (average power) of the output noise:

$$N = B_{out}(0) = A^2 W_0 B_s(0) = A^2 W_0 E \tag{12.30}$$

Let us set up the ratio of the peak value of the signal $s_{out}(t)$ to the root-mean-square value $\sqrt{N}$ of noise. In accordance with formulas (12.26) and (12.30),

$$s_{out}(t_0)/\sqrt{N} = AE/A\sqrt{W_0 E} = \sqrt{E/W_0} \tag{12.31}$$

Thus, in the case of white noise, the signal-to-noise ratio at the output of a filter matched to the signal depends only on the energy of the signal and the energy spectrum W_0 of noise.

From this it follows that, given the signal energy and spectrum width, the signal can be imparted any shape convenient for solving each particular problem.

Thus, for security of radio transmission, it is expedient to lengthen the signal, its amplitude being reduced correspondingly ($A_0^2 T_s = $ = const). This reduces the signal-to-noise ratio at the inputs of any receivers and makes it difficult to separate intelligence from the mixture of signals and noise. Only in a receiver matched to the given signal the maximum possible signal-to-noise ratio is restored. Of course, one must provide for the constant spectrum width when lengthening the signal. This can be achieved by using an internal pulse modulation, e.g., pulse frequency modulation [2]. An example of such a signal — the linear FM pulse — was considered in Sec. 3.7.

The lengthening of a pulse, supplemented by an internal pulse modulation, also makes it possible to reduce the peak power of the oscillator in the transmitter, the energy of the signal and its resolving power (after compression in the matched filter) remaining unchanged. This advantage is discussed in more detail in Sec. 12.5-2.

12.5. EXAMPLES OF CONSTRUCTING MATCHED FILTERS

12.5-1. Signal in the Form of a Square Video Pulse

Let a signal be defined by the following function of time:

$$s(t) = \begin{cases} 1 \text{ for } 0 \leqslant t \leqslant T_s \\ 0 \text{ for } t < 0 \text{ and } t > T_s \end{cases} \tag{12.32}$$

The spectral density of such a signal is known to be

$$S(\omega) = \frac{1}{i\omega}(1 - e^{-i\omega T_s}) \tag{12.33}$$

Using formula (12.16), in which t_0 is taken to be equal to the pulse duration T_s, we find the transfer function of the matched filter

$$\mathbf{K}(i\omega) = A\frac{1}{(-i\omega)}(1-e^{i\omega T_s})\,e^{-i\omega T_s} = A\frac{1-e^{-i\omega T_s}}{i\omega} \qquad (12.34)$$

A characteristic feature of this example is that $\mathbf{K}(i\omega)$ differs from the signal spectrum $\mathbf{S}(\omega)$ only in the constant coefficient. It is clear that the impulse response $g(t)$ of the filter coincides in shape with

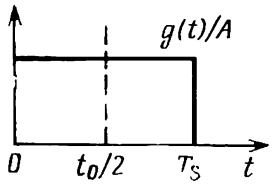


Fig. 12.6. Impulse response of a filter matched to a square pulse

the signal $s(t)$. In fact, from relation (12.22) it follows that

$$g(t) = As(T_s - t) = \begin{cases} A \text{ for } 0 \leqslant t \leqslant T_s \\ 0 \text{ for } t < 0 \text{ and } t > T_s \end{cases} \qquad (12.35)$$

The graph of $g(t)/A$ (Fig. 12.6) completely coincides in shape with the input pulse $s(t)$.

Further problem is reduced to finding the structure of a physical circuit having the impulse response shown in Fig. 12.6 and the transfer function according with formula (12.34).

A simplest signal defined by (12.32) is convenient for illustrating the main points of the synthesis of a two-port network, given its impulse response $g(t) = As(t_0 - t)$ or, which is the same, the complex transfer function $\mathbf{K}(i\omega)$ which is the Fourier transform of $g(t)$.

First of all, note that the integral

$$\int_0^\infty \frac{|\ln S(\omega)|}{1+\omega^2}\,d\omega = \int_0^\infty \frac{|\ln[2\sin(\omega T_s/2)] - \ln\omega|}{1+\omega^2}\,d\omega < \infty$$

i.e., the integral is convergent so that the function $K(\omega) = AS(\omega)$ does not contradict the Paley-Wiener criterion (12.23).

To study the possibility of realization of the required transfer function (12.34) by means of a two-port network with lumped parameters, let us consider the properties of this function on the p-plane:

$$\mathbf{K}(p) = A(1-e^{-pT_s})/p \qquad (12.36)$$

This function has no poles, since at the point $p = 0$, the numerator also vanishes; the number of zeros of the function $\mathbf{K}(p)$, which are the roots of the equation $1 - e^{-pT_s} = 0$ and are equal to $p_k = k(2\pi/T_s)$, is infinitely large. From this it follows that function (12.34) can be realized in a distributed-parameter system —

a length of line such that it takes a time T_s for the signal to run along it. The same conclusion can be easily arrived at by directly examining function (12.34). Obviously, the block diagram of this system must be such as shown in Fig. 12.7. The first factor $1/i\omega$ is realized

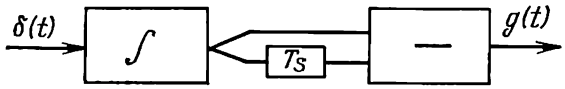


Fig. 12.7. Block diagram of a filter matched to a square pulse

by an integrating link and the second, $1 - e^{-i\omega T_s}$, by a subtractor which receives the signal both without delay and with a delay of T_s. The transfer function of an ideal (loss-free) delay line is equal to $e^{-i\omega T_s}$.

The operation of this circuit can also be explained by reasoning based on time notions: when a unit impulse [delta function $\delta(t)$] is applied to the input, a d-c voltage develops at the output of the

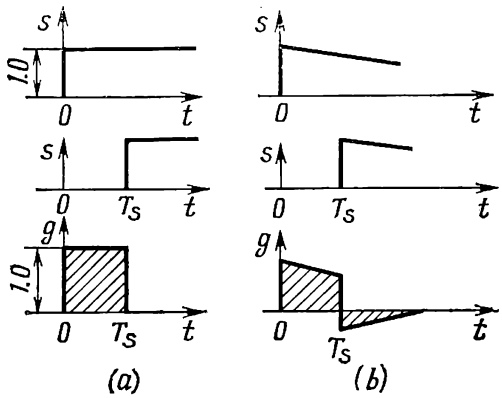


Fig. 12.8. Impulse response curves of (*a*) an ideal and (*b*) a physical filter

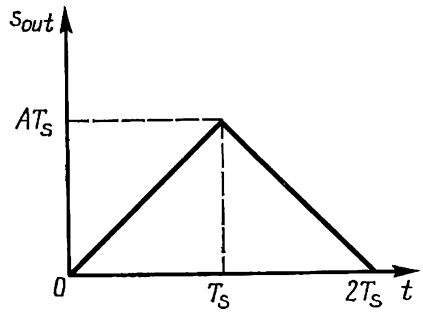


Fig. 12.9. Signal at the output of a filter matched to a square pulse

ideal integrator, starting from the moment $t = 0$. At the output of the device there will occur a voltage in the form of the difference between two unit steps shifted in time by an amount T_s with respect to each other (Fig. 12.8*a*).

The realization of the device shown in Fig. 12.7, which would provide for the accurate integration and delay of the input signal without distorting its shape (within the infinitely wide spectrum of the unit impulse) is practically impossible.

However, one can achieve a sufficiently good approximation by using a real integrating *R-C* circuit, provided its time constant is sufficiently large as compared with T_s. In this case, the voltage pulse at the subtractor output, which is the difference between two

exponential functions (Fig. 12.8*b*), can be made sufficiently close to rectangular.

Let us find the voltage at the filter output. Using formula (12.28) and taking into account that the autocorrelation function of the rectangular pulse has the shape of an isosceles triangle with a base $2T_s$ and an altitude equal to the pulse energy, we get

$$s_{out}(t) = AB^s(t-t_0) = \begin{cases} At & \text{for } 0 \leqslant t \leqslant T_s \\ A(2T_s - t) & \text{for } T_s \leqslant t \leqslant 2T_s \end{cases} \qquad (12.37)$$

(In the given case, for a unit amplitude, the pulse energy $E = T_s$.)

The output signal reaches its maximum value of AT_s at the moment $t = T_s$, i.e., by the end of the action of the input signal (Fig. 12.9).

In accordance with formula (12.31), the signal-to-noise ratio is

$$s_{out}(T_s)/\sqrt{N} = \sqrt{T_s/W_0} \qquad (12.38)$$

For comparison, let us set up a similar signal-to-noise ratio for an unmatched filter in the form of a single *R-C* link with a transfer function

$$\mathbf{K}(i\omega) = \frac{1/i\omega C}{R + 1/i\omega C} = \frac{1}{1 + i\omega RC}$$

The maximum value of the output signal, obtained at the moment $t = T_s$, is $s_{out}(T_s) = 1 - e^{-T_s/RC}$, while the root-mean-square value of noise is

$$\sigma_{out} = \sqrt{N} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} W_0 K^2(\omega)\, d\omega}$$

$$= \sqrt{\frac{W_0}{2\pi}\int_{-\infty}^{\infty} \frac{d\omega}{1 + (\omega RC)^2}} = \sqrt{\frac{W_0}{2RC}}$$

Thus,

$$\frac{s_{out}(T_s)}{\sqrt{N}} = \frac{1 - e^{-T_s/RC}}{\sqrt{W_0/2RC}} = \sqrt{\frac{T_s}{W_0}}\sqrt{\frac{2RC}{T_s}}\,(1 - e^{-T_s/RC})$$

At $T_s/RC = 1.28$ the product $\sqrt{2RC/T_s}\,(1 - e^{-T_s/RC})$ reaches its maximum value equal to about 0.77. Thus, the changeover from a single-link filter to a filter matched to the signal in this example increases the signal-to-noise ratio by a factor of $1/0.77 = 1.3$ (2.2 dB).

It will be clear from the examples that follow that the technical advantages of matched filtering manifest themselves essentially where signals are more complex.

12.5-2. Radio Pulse with a Frequency-Modulated Carrier

Let us consider the signal shown in Fig. 12.10*a*. The envelope of this signal has a rectangular shape, while the carrier frequency varies linearly (Fig. 12.10*b*) at a rate

$$\beta = 2\omega_d/T_s = 2 \times 2\pi f_d/T_s \tag{12.39}$$

where T_s is the pulse duration and $2\omega_d$ is the total frequency deviation within the pulse. The centre carrier frequency is $\omega_0 = 2\pi f_0$. Further we shall assume that $2\omega_d \ll \omega_0$.

Thus,

$$\omega(t) = \omega_0 + \beta t \tag{12.40}$$

and the instantaneous value of the signal within the interval from $-T_s/2$ to $T_s/2$ is defined by the expression

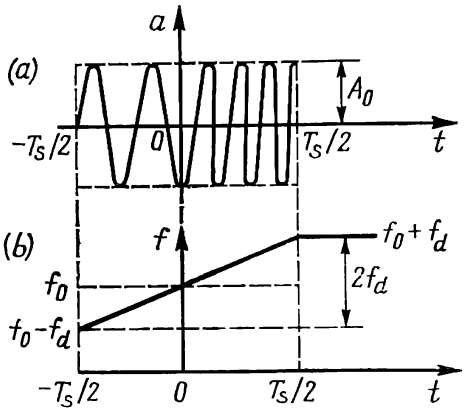


Fig. 12.10. (*a*) Linear FM pulse and (*b*) the law governing the variation of the instantaneous frequency

$$s(t) = A_0 \cos(\omega_0 t + \beta t^2/2) \tag{12.41}$$

The spectral density of such a signal was defined in Ch. 3. There it was found that the modulus and phase of the spectral density are defined by the following expressions [see (3.50) and (3.51)]:

$$S(\omega) = \frac{A_0 T_s}{2\sqrt{2m}} \sqrt{[C(u_1)+C(u_2)]^2 + [S(u_1)+S(u_2)]^2} \tag{12.42}$$

$$\varphi_s(\omega) = -\frac{\pi m}{4}\frac{(\omega-\omega_0)^2}{\omega_d^2} + \arctan\frac{S(u_1)+S(u_2)}{C(u_1)+C(u_2)} \tag{12.43}$$

In these expressions, $S(x)$ and $C(x)$ are the Fresnel integrals [see expression (3.49)], while u_1, u_2 and m are determined by the formulas

$$\left.\begin{aligned} u_1 &= \sqrt{\frac{\pi}{4}m}\left(1-\frac{\omega-\omega_0}{\omega_d}\right) \\ u_2 &= \sqrt{\frac{\pi}{4}m}\left(1+\frac{\omega-\omega_0}{\omega_d}\right), \quad m = 2f_d T_s \end{aligned}\right\} \tag{12.44}$$

Expressions (12.42) and (12.43) in principle can be used as a basis for synthesizing a filter. It is clear, however, that it is extremely difficult (if not entirely impossible) to develop a two-port network capable of an accurate realization of such complicated amplitude and frequency characteristics. One has, therefore, to employ various methods of approximation of the amplitude and frequency characte-

ristics. The first step is to assume that the envelope of the signal spectrum has a rectangular shape, while the phase characteristic has the shape of a quadratic parabola. Thus, precise expressions (12.42) and (12.43) are replaced by approximate ones [see the remarks to formulas (3.50) and (3.51)]:

$$S(\omega) \approx \frac{A_0 T_s}{2\sqrt{m}} = \text{const} \tag{12.45}$$

and

$$\varphi_s(\omega) \approx -\frac{\pi}{4} m \frac{(\omega-\omega_0)^2}{\omega_d^2} = -\frac{(\omega-\omega_0)^2}{2\beta} \tag{12.46}$$

for $\omega_0 - \omega_d < \omega < \omega_0 + \omega_d$.

In Sec. 3.7 it was shown that the higher the value of $m = 2f_d T_s$ the better the approximation (the constant phase shift $\pi/4$ is omitted).

When the time t is measured from the beginning of the pulse, the phase characteristic of the signal spectrum is written as

$$\varphi_s(\omega) = -\frac{\pi}{4} m \frac{(\omega-\omega)^2}{\omega_d^2} - \frac{\omega T_s}{2} \tag{12.47}$$

For a signal with such amplitude and phase spectra, the matched filter must have a rectangular amplitude-frequency characteristic and a phase-frequency characteristic determined by the expression

$$\varphi_K(\omega) = -\varphi_s(\omega) - \omega T_s = \left[\frac{\pi}{4} m \frac{(\omega-\omega_0)^2}{\omega_d^2} + \frac{\omega T_s}{2}\right]$$

$$-\omega T_s = \frac{\pi}{4} m \frac{(\omega-\omega_0)^2}{\omega_d^2} - \frac{\omega T_s}{2} \tag{12.48}$$

Strictly rectangular amplitude-frequency characteristic cannot practically be realized. Therefore, the next step towards simplification consists in the replacement of the rectangular frequency characteristic of the filter by an ordinary characteristic of a resonance filter. After that, the filter can practically be synthesized by combining the following two two-port networks: a band-pass resonance filter (an ordinary i-f amplifier of a radio receiver) and a special two-port network with a uniform amplitude characteristic and a square-law phase characteristic.

It should be noted that to the phase characteristic defined by (12.48) corresponds the delay

$$\tau(\omega) = \frac{d\varphi_K(\omega)}{d\omega} = \frac{\pi}{2} \frac{(\omega-\omega_0)}{\omega_d^2} - \frac{T_s}{2}$$

The device with the required phase characteristic can be any circuit whose delay in a certain frequency band (near ω_0) depends linearly on frequency. Such properties are characteristic of surface wave *dispersive* lines.

Let us determine the output signal of the "precise" matched filter whose transfer function satisfies condition (12.16).

Proceeding from relation (12.28), let us use expression (3.106′) for the autocorrelation function of the input signal, derived in Sec. 3.11:

$$B_a(\tau) \approx \frac{1}{2} A_0^2 T_s \frac{\sin[\pi m(\tau/T_s)(1-\tau/T_s)]}{\pi m(\tau/T_s)} \cos \omega_0 \tau$$

Substituting in this expression $t - T_s$ for τ and restricting the study to the vicinity of the point $t = T_s$, i.e., to the vicinity of the point where the output signal attains its peak value, we may consider that $t - T_s \ll T_s$.

Then

$$B_a(t-T_s) \approx \frac{1}{2} A_0^2 T_s \frac{\sin[(\pi m/T_s)(t-T_s)]}{(\pi m/T_s)(t-T_s)} \cos \omega_0 (t-T_s)$$

Taking into account that $m = (1/\pi)\,\omega_d T_s$ [see formula (3.38)], we write the latter expression in a modified form

$$B_a(t-T_s) = \frac{1}{2} A_0^2 T_s \frac{\sin \omega_d (t-T_s)}{\omega_d (t-T_s)} \cos \omega_0 (t-T_s) \qquad (12.49)$$

Substituting the obtained expression into (12.28), we find the output voltage of the matched filter:

$$v_{out}(t) = AB_a(t-T_s) = \frac{1}{2} AA_0^2 T_s \frac{\sin \omega_d (t-T_s)}{\omega_d (t-T_s)} \cos \omega_0 (t-T_s)$$
$$= V_{out}(t) \cos \omega_0 (t-T_s) \qquad (12.50)$$

where the envelope is

$$V_{out}(t) = \frac{1}{2} AA_0^2 T_s \frac{\sin \omega_d (t-T_s)}{\omega_d (t-T_s)} \qquad (12.51)$$

It should be noted that the carrier frequency is unmodulated and equal to ω_0, i.e., to the centre frequency of the input signal.

If the filter is lossless, the energy of the output signal is equal to the energy of the input signal. In this case, the following equality must be satisfied

$$\frac{1}{2} A_0^2 T_s = \frac{1}{2} \int_0^{2T_s} V_{out}^2(t)\,dt = \frac{1}{2} \int_0^{2T_s} \left(\frac{1}{2} AA_0^2 T_s\right)^2 \frac{\sin^2 \omega_d (t-T_s)}{[\omega_d (t-T_s)]^2}\,dt$$

or

$$\frac{1}{A^2} = \frac{A_0^2 T_s}{4\omega_d} \int_{-\omega_d T_s}^{\omega_d T_s} \frac{\sin^2 x}{x^2}\,dx$$

For high values of the parameter $m = 2f_d T_s$, the integration limits $\pm \omega_d T_s$ can be replaced by $\pm \infty$. In this case, the integral is equal to π, and finally

$$A = \frac{2}{A_0} \sqrt{\frac{\omega_d}{\pi T_s}} \qquad (12.52)$$

Substituting (12.52) into expression (12.51), we obtain

$$V_{out}(t) = \frac{1}{2}\left(\frac{2}{A_0}\sqrt{\frac{\omega_d}{\pi T_s}}\right) A_0^2 T_s \frac{\sin \omega_d (t-T_s)}{\omega_d (t-T_s)}$$

$$= \sqrt{m}\, A_0 \frac{\sin \omega_d (t-T_s)}{\omega_d (t-T_s)} \qquad (12.53)$$

The graphs of the input and output signals of the filter are given in Fig. 12.11 (for $A_0 = 1$). The highest amplitude of the output sig-

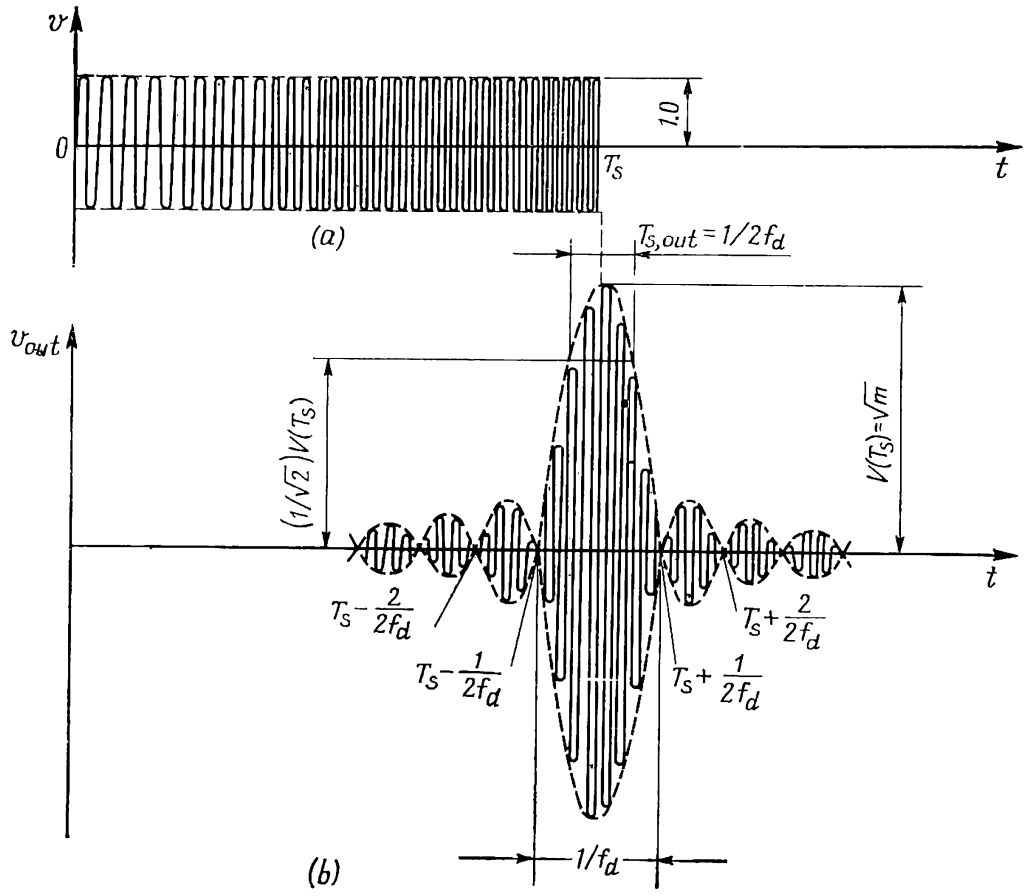


Fig. 12.11. (*a*) Linear FM pulse at the input of a matched filter and (*b*) the compressed signal at the output

nal (at the moment $t = T_s$) exceeds the amplitude of the input signal by a factor of $\sqrt{m}$ and the duration of the main lobe, measured between two zeros, is equal to $1/f_d$. When measured at the level of $1/\sqrt{2}$ of the maximum amplitude, the duration of the output pulse is approximately half that between two zeros, namely, $T_{s,\,out} \approx \approx 1/2\, f_d$.

Thus, the ratio

$$T_s/T_{s,out} \approx 2f_dT_s = m \tag{12.54}$$

coinciding in magnitude with the modulation parameter $m = 2f_dT_s$ may be called the *compression coefficient* of a frequency-modulated pulse in an optimum filter.

From expression (12.53) it is evident that the compensation of the phases of the signal spectrum, which constitutes the essence of matched filtering, leads in this case to a *decrease in the pulse duration by a factor of* m, *the signal amplitude being simultaneously increased by a factor of* $\sqrt{m}$. This property is very important in practice since it allows the pulse radiated by the transmitter to be lengthened for increasing the energy of the signal without any loss in the resolution which is defined by the duration of the output pulse of the matched filter [2]. The technical advantage of such a method manifests itself in cases where an increase in the pulse amplitude in the transmitter is limited by the pulse power of the electron devices used for generating the oscillations. It is much easier to raise the energy of the signal by lengthening the pulses and simultaneously applying frequency modulation. In this case, the modulation parameter m should rise in proportion to the duration T_s of the radiated signal (for a specified duration $T_{s,out}$ of the output pulse of the matched filter). In other words, the frequency deviation must remain unchanged, while the rate β of frequency change must be inversely proportional to the value of T_s (see Sec. 12.4).

12.5-3. Train of Identical Pulses

Let us consider a signal consisting of a group of n similar video pulses (Fig. 12.12). The intervals between the pulses may be different.

The spectrum of such a signal is

$$\mathbf{S}(\omega) = \mathbf{S}_1(\omega)(1 + e^{-i\omega T_1} + e^{-i\omega T_2} + \ldots + e^{-i\omega T_{n-1}}) \tag{12.55}$$

Here $\mathbf{S}_1(\omega)$ is the spectrum of the first pulse starting at the moment $t = 0$, $\mathbf{S}_1(\omega)e^{-i\omega T_1}$ is the spectrum of the second pulse starting at the moment $t = T_1$, etc.

Since the total duration of the signal shown in Fig. 12.12 is equal to $\tau_p + T_{n-1}$, then, in accordance with expression (12.16), the filter matched to the spectrum $\mathbf{S}(\omega)$ must have the transfer factor

$$\mathbf{K}(i\omega) = A\mathbf{S}^*(\omega)\, e^{-i\omega(\tau_p + T_{n-1})} = A\mathbf{S}_1^*(\omega)\, e^{-i\omega\tau_p} e^{-i\omega T_{n-1}} \times (1 + e^{i\omega T_1} + e^{i\omega T_2} + \ldots + e^{i\omega T_{n-1}})$$

$$= \mathbf{K}_1(i\omega)\,[1 + e^{-i\omega(T_{n-1}-T_{n-2})} + e^{-i\omega(T_{n-1}-T_{n-3})} + \ldots + e^{-i\omega(T_{n-1}-T_1)} + e^{-i\omega T_{n-1}}] \quad (12.56)$$

In this expression, $\mathbf{K}_1(i\omega) = A\mathbf{S}_1^*(\omega)e^{-i\omega\tau_p}$ is the transfer factor of the filter matched to a single impulse.

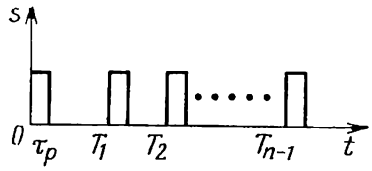


Fig. 12.12. Signal in the form of a train of pulses

On the basis of expression 12.56 one can easily sketch a diagram of the filter matched to the signal shown in Fig. 12.12. Such a filter must contain a link with a transfer function $\mathbf{K}_1(i\omega)$ for the optimum intrapulse processing of the signal, and a set of delay lines. The magnitudes of the delays must increase in the reverse order with respect to the sequence of pulses in the train at the filter input. One of the possible versions of such a device is shown in Fig. 12.13*a*.

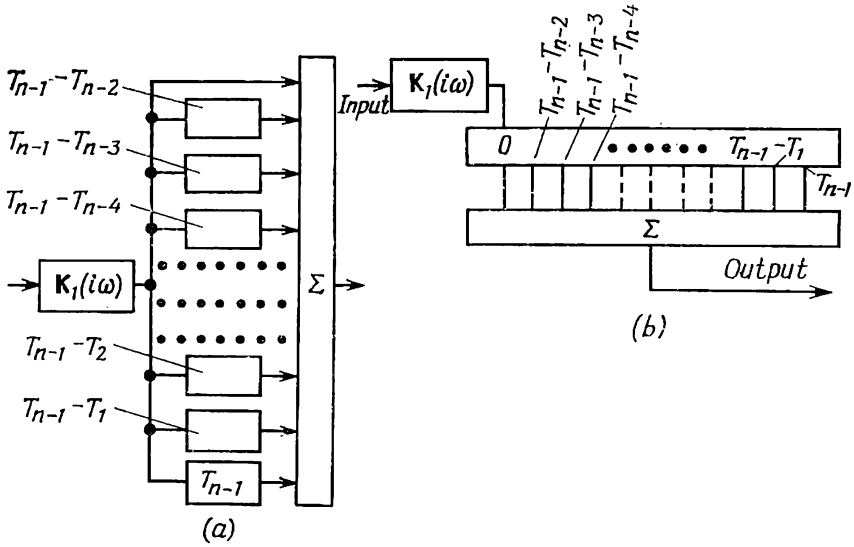


Fig. 12.13. Matched filtering of a train of pulses (to Fig. 12.12)

The maximum pulse is obtained at the adder output when the first pulse of the input sequence passing through the delay line T_{n-1} is added to the second pulse passing through the delay line $T_{n-1} - T_1$, to the third pulse delayed by $T_{n-1} - T_2$, etc. up to the last pulse passing through this device without any additional delay.

Instead of a set of n delay lines, it is more simple and expedient to employ a single delay line with n taps, as shown in Fig. 12.13*b*. The taps are arranged in such a way that the corresponding delays increase in the same order as in Fig. 12.13*a*.

The structure of the matched filter is much simplified when the input signal is a train of equidistant pulses, i.e., when

$$T_1 = T,\ T_2 = 2T,\ T_3 = 3T,\ \ldots\ T_{n-1} = (n-1)\,T$$

For this case, expression (13.56) can be written as

$$\mathbf{K}(i\omega) = \mathbf{K}_1(i\omega)\,[1 + e^{-i\omega T} + e^{-i2\omega T} + \ldots + e^{-i(n-1)\omega T}] = \mathbf{K}_1(i\omega)\,\mathbf{K}_2(i\omega) \quad (12.57)$$

With sufficiently large n, the expression in the brackets can be transformed using the formula of geometric progression:

$$\mathbf{K}_2(i\omega) = 1/(1 - e^{-i\omega T}) \quad (12.58)$$

The structure of expression (12.57) indicates to a possibility of realization of a matched filter in the form of two cascade-connected two-port networks: one with a transfer function $\mathbf{K}_1(i\omega)$, matched to a single impulse as in the circuit shown in Fig. 12.13*a*, and the other in the form of a circuit with negative feedback containing only one delay line T (Fig. 12.14). The transfer function of such a circuit (which is enclosed in a dashed box in Fig. 12.14) is determined by the expression

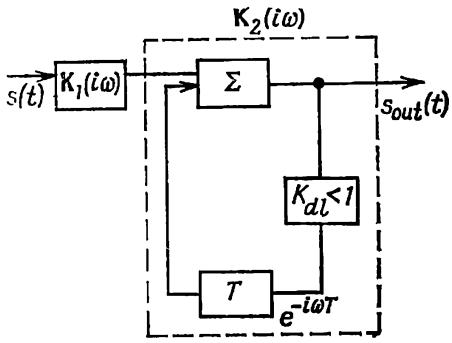


Fig. 12.14. Comb filter

$$\mathbf{K}_2(i\omega) = \frac{1}{1 - K_{dl}e^{-i\omega T}} = 1 + K_{dl}e^{-i\omega T} + K_{dl}^2 e^{-i2\omega T} + \ldots \quad (12.59)$$

The symbol K_{dl} stands for a lag-free two-port network taking account of the attenuation (amounting to tens of decibels) in the delay line and including an amplifier compensating for this attenuation. To provide for the stability of the circuit, the coefficient K_{dl} must be less than unity. In this case, the delay line may be regarded as an ideal line with a transfer function $e^{-i\omega T}$.

For frequencies meeting the requirement $\omega T = (2k + 1)\,\pi$, $k = 0, 1, 2, \ldots$, feedback is negative and $\mathbf{K}_2(i\omega) = 1/(1 + K_{dl})$. For frequencies $\omega T = 2k\pi$, $k = 0, 1, 2, \ldots$, feedback is positive and $\mathbf{K}_2(i\omega) = 1/(1 - K_{dl})$.

The amplitude-frequency characteristic of the circuit assumes the form shown in Fig. 12.15. The filters having such a characteristic are called *comb* filters. They are effective for separation of signals

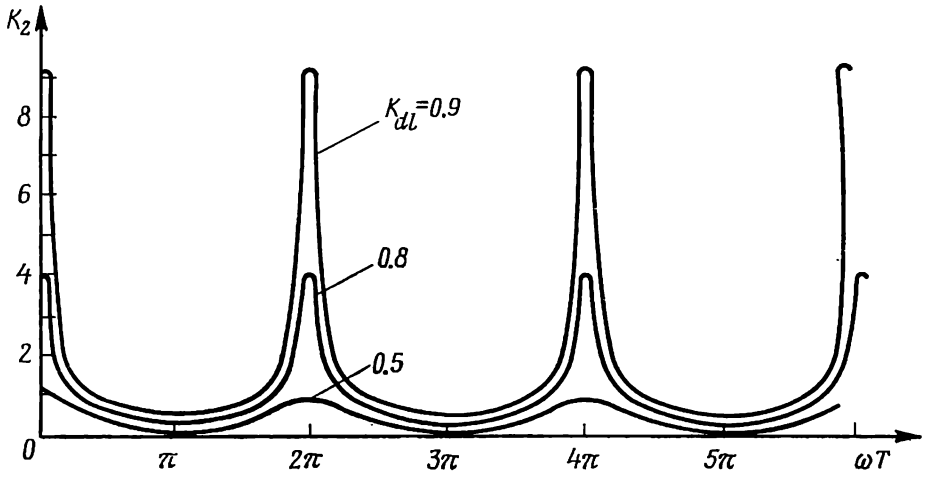


Fig. 12.15. Amplitude-frequency characteristic of a comb filter

in the form of a periodic train of pulses on the background of white noise. The greater the number n of the pulses in a train and the closer the coefficient K_{dl} to unity, the closer the circuit approximates the matched filter.

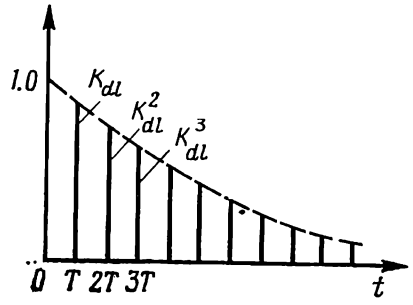


Fig. 12.16. Impulse response of a comb filter

The impulse response $\mathbf{K}_2(i\omega)$ of the filter is defined by the obvious expression

$$g(t) = \delta(t) + K_{dl}\delta(t - T) + K_{dl}^2\delta(t - 2T) + \ldots$$

The coefficients multiplying the unit impulses that arise at the output of the two-port network at intervals of T decay according to a law close to exponential (for K_{dl} close to unity). Thus, the impulse response K_2 of the filter has the shape shown in Fig. 12.16 (the delta functions are not shown in Fig. 12.16).

All the above arguments are applicable to the filtering of a train of radio pulses. But in this case, by $\mathbf{K}_1(i\omega)$ one must mean the transfer factor of the filter matched to a single impulse of the input train.

12.6. SHAPING OF THE SIGNAL WHICH IS THE COMPLEX CONJUGATE OF A GIVEN SIGNAL

Let us consider an interesting property of the circuit shown in Fig. 12.17. In this circuit, $\mathbf{K}_1(i\omega)$ and $\mathbf{K}_2(i\omega)$ are the transfer functions of the filters at the receiving and transmitting ends of the

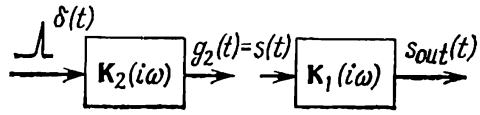


Fig. 12.17. Shaping a signal matched to a given filter

communication channel, and

$$\mathbf{K}_1(i\omega) = \mathbf{K}_2^*(i\omega)$$

i.e., the functions $\mathbf{K}_1(i\omega)$ and $\mathbf{K}_2(i\omega)$ are the complex conjugates.

When the two-port network $\mathbf{K}_2(i\omega)$ is shock excited by a unit impulse $\delta(t)$, at the output there develops oscillation (impulse response)

$$g_2(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{K}_2(i\omega)\, e^{i\omega t}\, d\omega$$

which is used as the signal to be transmitted through the communication channel. Thus, $g_2(t) = s(t)$.

It is evident that the receiving filter $\mathbf{K}_1(i\omega)$ is matched to this signal since its impulse response $g_1(t)$ is a mirror image of the signal $s(t)$.

Indeed,

$$g_1(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{K}_1(i\omega)\, e^{i\omega t}\, d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{K}_2^*(i\omega)\, e^{i\omega t}\, d\omega$$

$$= g_2(-t) = s(-t)$$

[The constant delay t_0 entering into expression (12.22) is omitted here.]

The signal $s_{out}(t)$ at the output of the filter $\mathbf{K}_1(i\omega)$ is maximized in the sense of expression (12.17).

Thus, the signal matched to a specified receiving filter can be shaped at the transmitting end by shock-exciting an "inverted" filter by a unit impulse. The "inverted" filter is a filter whose transfer function is the complex conjugate of the transfer function of the "direct" filter.

Since the shaping of signals and their processing in the receiver are ordinarily effected at the intermediate frequency, the system illustrated by Fig. 12.17 must be complemented with a high-frequency oscillator and a converter for shifting the signal spectrum to

the high frequency region in the transmitter, and with a local oscillator and a converter for inverting the frequency in the receiver.

In spite of the apparent simplicity of this principle of shaping a signal that provides for its optimal processing in a radio receiver, the realization of the inverted filter is a complex problem which can be solved successfully not for any signal.

This problem is relatively simple in the case of a communication channel using a phase-shift-keyed signal which is a train of radio

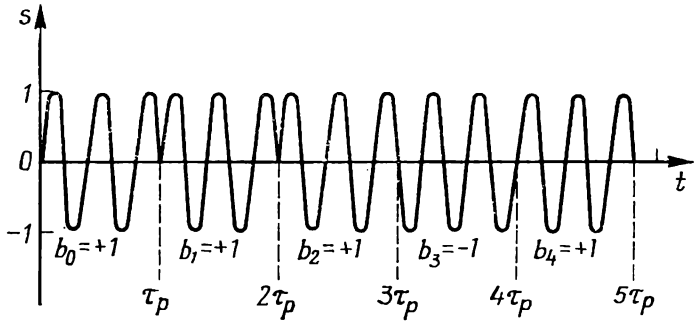


Fig. 12.18. Phase-shift-keyed radio-frequency oscillation

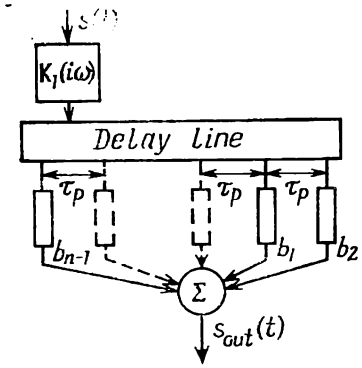


Fig. 12.19. Block diagram of a filter matched to a phase-shift-keyed signal

pulses following one another without intervals and distinguished only by the phase of the r-f carrier. The epoch angle in each radio pulse can be either 0 or π, the alternation of the phase angles following a definite code: the kth pulse is ascribed a coefficient b_k equal to ± 1. The plus sign corresponds to the phase angle 0 and the minus sign, to the phase angle π.

Figure 12.18 shows such a signal of five radio pulses with coefficients $b_0 = +1$, $b_1 = +1$, $b_2 = +1$, $b_3 = -1$, and $b_4 = +1$.

A block diagram of a filter used for processing such a signal is shown in Fig. 12.19. The filter is a combination of a two-port network $K_1(i\omega)$ matched to a single impulse (of duration τ_p) and a multitap delay line. The number of the taps following at intervals of τ_p

is equal to the number of the elementary radio pulses in the signal. The lag-free two-port networks b_0, b_1, b_2, . . . let pass through the pulses fed from the taps of the delay line, the phase of the r-f carrier

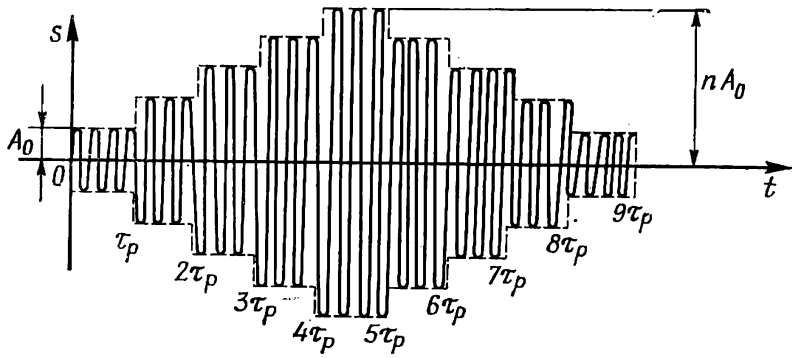


Fig. 12.20. Oscillation at the output of a filter matched to a phase-shift-keyed signal

being either left unchanged or shifted by 180°.

The alternation of $b_0, b_1, \ldots$ follows a pattern that is the mirror image of the signal. As a result, the output voltage has the shape shown in Fig. 12.20 (without taking into account the effect of the two-port network $\mathbf{K}_1(i\omega)$ on the pulse shape).

By the end of the action of the input signal, a maximum pulse with an amplitude nA_0 (where n is the number of the elementary pulses) is produced at the adder output. Thus, this circuit effects the *compression of the signal*, the compression ratio being equal to n, i.e., to the number of the taps in the delay line. In this case, the number n plays the same role as the product $2f_dT_s = m$ for a filter effecting the compression of radio pulses with a frequency-modulated carrier [see formula (12.44)].

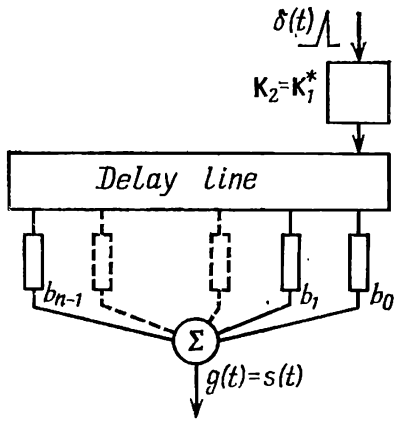


Fig. 12.21. Block diagram of a filter which is the inverse of the filter shown in Fig. 12.19

A block diagram of the inverted filter for shaping the signal shown in Fig. 12.18 is illustrated in Fig. 12.21. This diagram differs from that shown in Fig. 12.19 in that the input signal is applied to the opposite end of the delay line, so that the alternation of the coefficients b_0, b_1, . . ., b_{n-1} follows a pattern which is the mirror image of that in the circuit shown in Fig. 12.19. Furthermore, the transfer function $\mathbf{K}_2(i\omega)$ of the two-port network performing the intrapulse

processing is the complex conjugate of the function $\mathbf{K}_1(i\omega)$ shown in Fig. 12.19. For a pulse that is symmetrical with respect to its centre, $\mathbf{K}_2(i\omega)$ coincides with $\mathbf{K}_1(i\omega)$. In fact, the filters shown in Figs. 12.19 and 12.21. are quite identical and this is a great advantage, particularly in cases where the receiver and the transmitter are located in one place, e.g., in a radar system. In such cases, the generation of a signal and its optimum processing in reception can be effected by means of a single filter. Such a system might be called *key-lock system*.

12.7. MATCHED FILTERING OF A SPECIFIED SIGNAL IN THE CASE OF NONWHITE NOISE

Assume that noise with a nonuniform energy spectrum $W(\omega)$ (nonwhite noise) is linearly (additively) superimposed on a completely known signal $s(t)$. It is necessary to synthesize a filter maximizing the signal-to-noise ratio. In contrast to the previously discussed

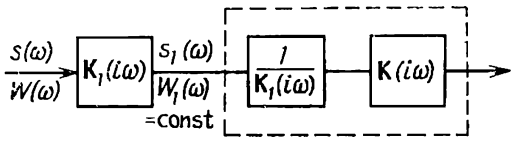


Fig. 12.22. Matched filtering of a specified signal contaminated with nonwhite noise

problems, in this case the transfer function must be matched both to the signal spectrum $\mathbf{S}(\omega)$ and to the energy spectrum $W(\omega)$ of noise.

A most simple method of finding the required transfer function $\mathbf{K}(i\omega)$ is to transform the given noise into white noise [1]. To reveal the essence of this method, let us consider an auxiliary block diagram shown in Fig. 12.22. In this diagram, $\mathbf{K}(i\omega)$ stands for the sought for transfer function of the filter to be synthesized, while $\mathbf{K}_1(i\omega)$ and $1/\mathbf{K}_1(i\omega)$ are the transfer functions of two auxiliary, conditional two-port networks which have no effect on the operation of the device, because their resultant transfer function is equal to unity.

Since the function $\mathbf{K}_1(i\omega)$ can be selected arbitrarily, let us specify the modulus of this function in the form

$$K_1(\omega) = \sqrt{W_0/W(\omega)} \tag{12.60}$$

where W_0 is a constant.

In this case at the output of the first two-port network there will occur noise with a uniform energy spectrum

$$W_1(\omega) = W(\omega)\,[K_1(\omega)]^2 = W_0 = \text{const}$$

i.e., *white noise*.

Of course, the signal at the output of this two-port network differs from the input signal, because the spectral density

$$\mathbf{S}_1(\omega) = \mathbf{S}(\omega)\,\mathbf{K}_1(i\omega) \tag{12.61}$$

differs from $\mathbf{S}(\omega)$. However, this is not important; the main task is to provide the maximum signal-to-noise ratio at the output of the entire system. What is important is the ratio of the signal energy to the energy spectrum of noise, whereas the shape of the signal in this case is of no concern.

Since in the given section of the system noise is white, then, to obtain the maximum signal-to-noise ratio at the output, the transfer function of the entire subsequent part of the system must satisfy condition (12.16). Thus, we get

$$\frac{1}{\mathbf{K}_1(i\omega)}\,\mathbf{K}(i\omega) = A\mathbf{S}_1^*(\omega)\,e^{-i\omega t_0} \tag{12.62}$$

The left-hand side of this expression is the resultant transfer function of the two-port network enclosed in a dashed box in Fig. 12.22, while the right-hand side is the complex conjugate of the spectrum $\mathbf{S}_1(\omega)$, supplemented by the factor $e^{-i\omega t_0}$

From expression (12.62) we obtain

$$\mathbf{K}(i\omega) = A\mathbf{S}_1^*(\omega)\,\mathbf{K}_1(i\omega)\,e^{-i\omega t_0} \tag{12.63}$$

However, from (12.61) it follows that

$$\mathbf{S}_1^*(\omega) = \mathbf{S}^*(\omega)\,\mathbf{K}_1^*(i\omega)$$

Thus,

$$\mathbf{K}(i\omega) = A\mathbf{S}^*(\omega)\,\mathbf{K}_1^*(i\omega)\,\mathbf{K}_1(i\omega)\,e^{-i\omega t_0} = A\mathbf{S}^*(\omega)\,[K_1(\omega)]^2\,e^{-i\omega t_0}$$

Substituting relation (12.60) into the above expression, we obtain

$$\mathbf{K}(i\omega) = AW_0\,\frac{\mathbf{S}^*(\omega)}{W(\omega)}\,e^{-i\omega t_0} \tag{12.64}$$

The physical meaning of this relation can easily be interpreted.

As in the case of white noise, to maximize the signal-to-noise ratio, the filter must effect compensation of the epoch angles of the spectrum $\mathbf{S}(\omega)$ of the input signal. Therefore, the right-hand side of expression (12.64) includes the complex conjugate $\mathbf{S}^*(\omega)$. However, the modulus of the transfer function must, in the first place, be proportional to the modulus of $\mathbf{S}(\omega)$ (as in the case of white noise) and, in the second place, be inversely proportional to the energy spectrum of noise at the filter input. In this way, those components of the signal spectrum are accentuated, for which the noise intensity is less.

12.8. FILTERING OF A SIGNAL OF UNKNOWN EPOCH ANGLE

When processing complex signals with intrapulse modulation, the epoch angle θ_0 of the r-f carrier in the expression

$$a(t, \theta_0) = A(t)\cos[\omega_0 t + \theta(t) + \theta_0] \tag{12.65}$$

is usually unknown.

If one has a filter that is matched to a signal $a(t, 0)=A(t)\times$ $\times\cos[\omega_0 t+\theta_i(t)]$ without taking θ_0 into account, in the presence of the phase shift θ_0 the filter becomes mismatched. Let us reveal the effect of this mismatch on the output oscillation.

Using general expression (12.28), let us represent the output signal of the filter (first for $\theta_0 = 0$) in the form of (3.98), (3.99):

$$a_{out}(t\ 0) = \frac{A}{2}\operatorname{Re}\left[e^{i\omega_0 t}\int_{-\infty}^{\infty}\mathbf{A}(x)\,\mathbf{A}^*(x-t)\,dx\right]$$

$$=\frac{1}{2}A\ \operatorname{Re}[e^{i\omega_0 t}B_A(t)] \qquad (12.66)$$

To simplify the analysis, the constant delay t_0 entering into (12.28) is omitted. Therefore, the shift τ is replaced by t. Furthermore, the minus sign in the exponent is replaced by the plus sign in accordance with the sense of the shift of the function $\mathbf{A}^*(x-t)$. As mentioned in Sec. 3.10, the integral in expression (12.66) has the meaning of the autocorrelation function $B_A(t)$ of the complex envelope $\mathbf{A}(t)$.

Let us introduce into consideration the epoch angle θ_0 of the input signal. For this purpose, it is sufficient to multiply the function $\mathbf{A}(x)$ by $e^{i\theta_0}$. The new integral

$$\int_{-\infty}^{\infty}\mathbf{A}(x)\,e^{i\theta_0}\mathbf{A}^*(x-t)\,dx \qquad (12.67)$$

defines the cross-correlation between the functions $\mathbf{A}(x)e^{i\theta_0}$ and $\mathbf{A}^*(x-t)$; however, after taking the factor $e^{i\theta_0}$ outside the integral sign, we obtain the product $e^{i\theta_0}B_A(t)$.

Thus we come to the following expression for the signal at the output of a *mismatched filter*:

$$a_{out}(t, \theta_0) = \frac{A}{2}\operatorname{Re}[e^{i(\omega_0 t+\theta_0)}B_A(t)] \qquad (12.68)$$

From comparison of this expression with (12.66) it follows that to take account of the epoch angle, it is sufficient to add θ_0 to the term $\omega_0 t$ while preserving the envelope of the output signal.

Let us illustrate this result with an example of a linear FM pulse considered in Secs. 3.11 and 12.5.

From relations (3.103) and (3.106′), in which τ is replaced by t (the signal delay is not taken into account), stems the following expression for the autocorrelation function of the envelope:

$$B_A(t) = T_s\frac{\sin\left[\pi m\frac{t}{T_s}\left(1-\frac{t}{T_s}\right)\right]}{\pi m\,(t/T_s)}$$

Thus,

$$a_{out}(t, \theta_0) = AT_s \frac{\sin\left[\pi m \frac{t}{T_s}\left(1-\frac{t}{T_s}\right)\right]}{\pi m (t/T_s)} \cos(\omega_0 t + \theta_0) \qquad (12.69)$$

Figure 12.23 shows the output oscillation in a section of the time axis near the peak at $\theta_0 = 90°$ for a filter matched to the linear FM signal. The parameters of the input signal correspond to those given in Sec. 12.5-2 (Fig. 12.10). Depending on the value of θ_0, the position of the peak of the compressed signal on the time axis can vary within $\pm \pi/\omega_0$, i.e., within half a period of the r-f carrier. From this example it is obvious that with the number of periods per duration of the compressed signal being sufficiently large, the influence of θ_0 on the height of the peak is insignificant. If further processing of the signal is effected using the envelope, the effect of θ_0 is eliminated provided the above condition as to the r-f carrier is satisfied.

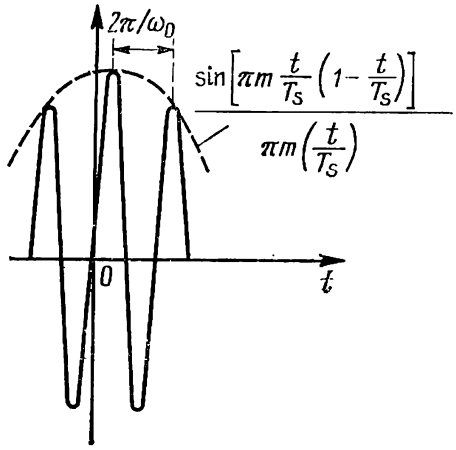


Fig. 12.23. Radio-frequency carrier of a linear FM pulse with the epoch angle θ_0 of the signal equal to 90°

12.9. MATCHED FILTERING OF A COMPLEX SIGNAL

In Chaps. 3, 6 and others it was noted that the complex envelope $\mathbf{A}(t)$ of a narrow-band signal $a(t) = A(t) \cos[\omega_0 t + \theta(t)]$ contains all the information imposed on the carrier by both amplitude and angle modulation. In many practical cases the signal processing is preferably effected directly using the envelope $\mathbf{A}(t)$, the carrier frequency ω_0 being excluded.

A block diagram of a device for the separation of the complex envelope of a narrow-band signal $a(t)$ is shown in Fig. 12.24. The device consists of two identical frequency converters with a common local oscillator whose frequency ω_0 coincides with the carrier frequency of the signal.

The selective circuit at the output of each converter is a low-pass filter (R-C circuit). The passband of the filter is assumed to be sufficient for undistorted reproduction of the spectrum of the intelligence being transmitted. When the condition $E_{lo} \gg A_{max}$ is satisfied, linear frequency conversion is effected, so that the oscillation at the output of the first converter assumes the form [see formulas

(8.74) and (8.77)]

$$A_{\cos}(t) = a_2 E_{lo} A(t) \cos\{[\omega_0 t + \theta(t)] - \omega_0 t\} = k_{con} A(t) \cos\theta(t) \quad (12.70)$$

Because of the shift of the phase of the local oscillator e.m.f. by an angle $\varphi = 90°$, the oscillation at the output of the second converter is

$$A_{\sin}(t) = a_2 E_{lo} A(t) \cos\{[\omega_0 t + \theta(t)] - (\omega_0 t + 90°)\} = k_{con} A(t) \sin\theta(t) \quad (12.71)$$

The symbol $k_{con} = a_2 E_{lo}$ stands for the constant coefficient having the sense of the slope of the characteristic curve of the converter, and a_2 is the coefficient multiplying the quadratic term in expression (8.10).

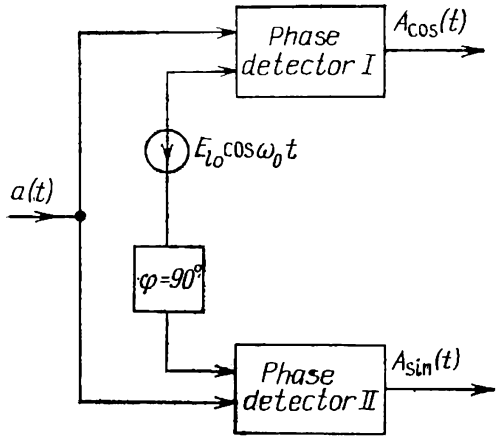


Fig. 12.24. Separation of the quadrature components of the complex envelope of a narrow-band oscillation

Converters producing an output oscillation containing information on the phase $\theta(t)$ (the term $\omega_0 t$ is excluded) are usually called *phase detectors.*

The cosine oscillation $A_{\cos}(t)$ and the sine oscillation $A_{\sin}(t)$ coincide respectively with the real and imaginary parts of the complex envelope $\mathbf{A}(t)$ [see expression (3.88)]. In this sense the signal processing in question is a *quadrature process.*

The combination of the physical oscillations $A_{\cos}(t)$ and $A_{\sin}(t)$, written in the form of the sum $A_{\cos}(t) + iA_{\sin}(t)$, makes it possible to treat the complex oscillation as a physical process, but it should be borne in mind that the complex oscillation under consideration is not an analytic signal. This is due to the fact that the spectral density of the complex envelope $\mathbf{A}(t)$ does not vanish in the frequency region $\omega < 0$ (see Para. 3 of Sec. 3.10).

Further processing of a complex oscillation should be carried out in accord with the common rules for operations with complex numbers.

If a narrow-band signal

$$a(t) = A(t) \cos[\omega_0 t + \theta(t)] = \mathrm{Re}[\mathbf{A}(t) e^{i\omega_0 t}]$$

after passing through a filter with an impulse response

$$g(t) = G(t) \cos[\omega_0 t + \gamma(t)] = \mathrm{Re}[\mathbf{G}(t) e^{i\omega_0 t}] \quad (12.72)$$

assumes the form (see Sec. 6.6)

$$a_{out}(t) = A_{out}(t)\cos[\omega_0 t + \theta_{out}(t)] = \mathrm{Re}[\mathbf{A}_{out}(t)\, e^{i\omega_0 t}]$$

then, after equivalent quadrature processing, this signal must have the following form at the output of the low-pass filter:

$$\mathbf{A}_{out}(t) = \int_{-\infty}^{\infty} \mathbf{A}(x)\,\mathbf{G}(t-x)\,dx \qquad (12.73)$$

From this requirement it follows that the impulse response of the filter sought for must coincide with the *complex* function $\mathbf{G}(t)$. The complex impulse response can be realized by means of two physical filters with real impulse characteristics $G_{\cos}(t) = G(t)\cos\gamma(t)$ and $G_{\sin}(t) = G(t)\sin\gamma(t)$, respectively [see formula (12.72)]:

$$\mathbf{G}(t) = G_{\cos}(t) + iG_{\sin}(t) \qquad (12.74)$$

Substituting $\mathbf{A}(t)$, defined by formulas (12.70) and (12.71), and $\mathbf{G}(t)$, defined by formula (12.74), into (12.73), we get

$$\mathbf{A}_{out}(t) = \int_{-\infty}^{\infty} [A_{\cos}(x) + iA_{\sin}(x)]\,[G_{\cos}(t-x) + iG_{\sin}(t-x)]\,dx$$

$$= \int_{-\infty}^{\infty} A_{\cos}(x)\,G_{\cos}(t-x)\,dx - \int_{-\infty}^{\infty} A_{\sin}(x)\,G_{\sin}(t-x)\,dx$$

$$+ i\left[\int_{-\infty}^{\infty} A_{\sin}(x)\,G_{\cos}(t-x)\,dx + \int_{-\infty}^{\infty} A_{\cos}(x)\,G_{\sin}(t-x)\,dx\right] \qquad (12.75)$$

The first integral defines the response of the physical filter having the real impulse response $G_{\cos}(t)$ to the stimulus $A_{\cos}(t)$, the second integral defines the response of the filter having the impulse response $G_{\sin}(t)$ to the stimulus $A_{\sin}(t)$, etc.

Algorithm (12.75) is realized in the circuit shown in Fig. 12.25. This is a general algorithm applicable to an arbitrary function $\mathbf{G}(t)$, but in the case of *matched* filtering, the impulse response $\mathbf{G}(t)$ must satisfy the condition stemming from the requirement of maximizing the peak of the signal. This condition, by analogy with expression (12.22) and with account taken of (12.66), can be written in the form

$$\mathbf{G}(t) = C\mathbf{A}^*(t)$$

where C is a constant coefficient.

Thus,

$$\mathbf{G}(t) = C\mathbf{A}^*(t) = C[A_{\cos}(t) - iA_{\sin}(t)] \qquad (12.76)$$

Substitution of this expression into (12.75) yields the following result:

$$\mathbf{A}_{out}(t) = C \int_{-\infty}^{\infty} \mathbf{A}(x)\,\mathbf{A}^*(x-t)\,dx = CB_A(t) \qquad (12.77)$$

From this expression it follows that in the case of strictly matched filtering the signal at the output of the system under consideration is a real function coinciding in shape with the autocorrelation function of the complex envelope of the input oscillation.

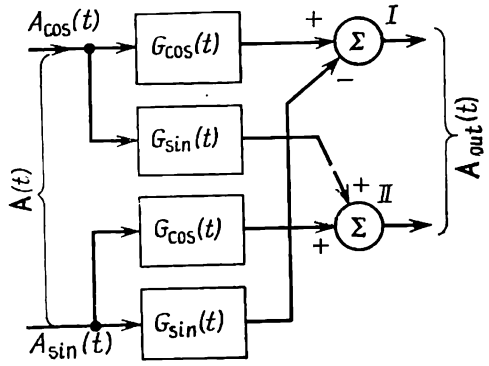


Fig. 12.25. Complex filter

Let us introduce into consideration the unknown epoch angle θ_0. To this end, the functions $A_{\cos}(x)$ and $A_{\sin}(x)$ entering into expression (12.75) must be multiplied by $e^{i\theta_0}$ to obtain

$$\mathbf{A}_{out}(t) = CB_A(t)\,e^{i\theta_0} = CB_A(t)\cos\theta_0 + iCB_A(t)\sin\theta_0 \qquad (12.78)$$

For $\theta_0 \neq 0$, the resultant oscillation at the output of the adders I and II is complex. Figure 12.26 illustrates (to the left of the dashed

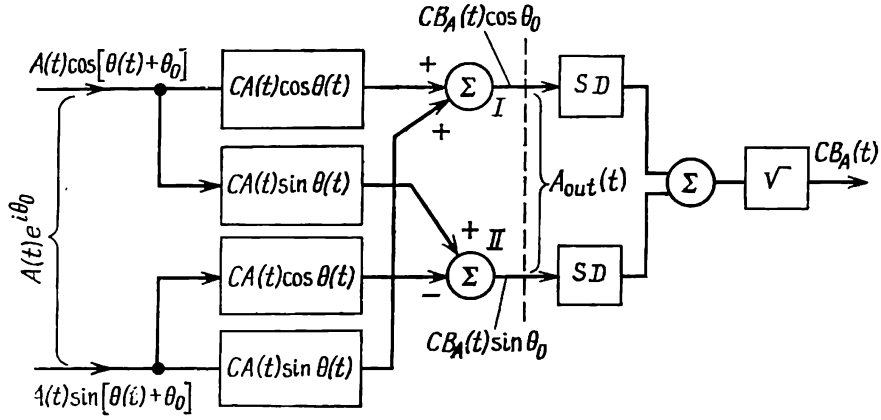


Fig. 12.26. Complex filter matched to the complex envelope of a narrow-band oscillation and supplemented with elements for square-law processing

line) the circuit shown in Fig. 12.25, but with changed polarities of the sine and cosine filters. The impulse responses of these filters remain the same. Shown to the right of the dashed line are addition-

al devices needed for the elimination of the unknown epoch angle: squaring devices (*SD*), an adder, and a device for the square rooting of the sum of squares. This additional processing yields an oscillation $CB_A(t)$ coinciding with (12.77).

Continuing with the example of the matched filtering of a linear FM pulse given in the preceding section, we obtain

$$\mathbf{A}_{out}(t) = CT_s e^{i\theta_0} [\sin \pi m (t/T_s)(1 - t/T_s)]/\pi m(t/T_s)$$

and after complete processing according to the diagram shown in Fig. 12.26, we have

$$A_{out}(t) = CT_s [\sin \pi m (t/T_s)(1 - t/T_s)]/\pi m (t/T_s)$$

Squaring and square rooting are nonlinear conversions of the signal. However, these processes take place after the maximization of the signal-to-noise ratio, therefore the effect of interaction of the signal and noise is less harmful than in the case of the direct action of a signal and noise on a nonlinear device.

Chapter 13

DISCRETE PROCESSING OF SIGNALS. DIGITAL FILTERS

13.1. GENERAL

Recent years have been characterized by the fast development of discrete control systems and data transmission systems making wide use of mathematical modelling of filtering processes, based on the application of digital computers. This new direction in radioelectronics has a great effect on the development of the theory of circuits and signals and its practical applications.

A general idea of the principle of digital processing of a continuous signal can be got from the diagram shown in Fig. 13.1. The same

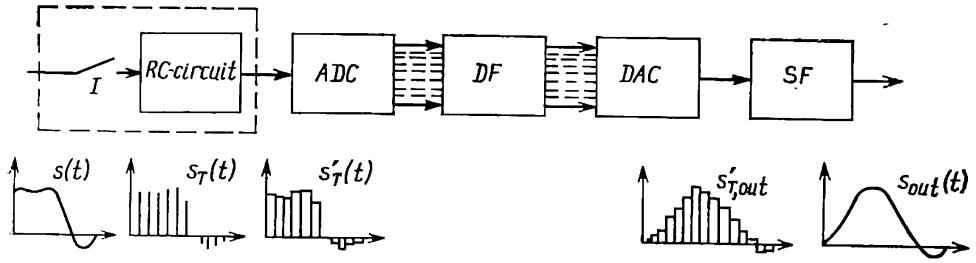


Fig. 13.1. Functional diagram of a digital filter

figure shows the graphs of oscillations in different sections of the circuit.

The input signal $s(t)$ is first subjected to time quantization, or sampling, effected by means of an electronic switch operating so that samples are taken at regular intervals T. The sampled signal $s_T(t)$ at the output of the electronic switch (ES) has the form of short equidistant pulses which are samples of the signal $s(t)$. Each sample is stored in an integrating R-C circuit for a time required for the operation of an analog-to-digital converter (ADC). The storage time must be shorter than the interval T. As a result, at the output of the R-C circuit there occurs a stepped oscillation $s'_T(t)$.

In the ADC each sample is subjected to level quantization and is converted into a coded word — a binary number of r digits each of which is represented by a zero or a unity (pause or standard pulse).

The quantization consists in that a sample is measured and is assigned a single level from a number of possible levels. This num-

ber is equal to 2^r. For example, for $r = 10$, there are $2^{10} = 1024$ levels. Each digit is transmitted through its own bus so that the coded sample at the *ADC* output is represented as a combination of binary units (pauses and pulses) appearing simultaneously on r output buses (parallel code). To the maximum possible value of the sample corresponds a coded word of r pulses, and to its zero value, that of r pauses. The longer the coded word, i.e., the greater the number of the binary units in this word, the higher the accuracy of representation.

A sequence of coded samples is fed into a digital filter (*DF*) comprising a computer in which the coded words are subjected to mathematical operations (addition, subtraction, multiplication, and delaying) corresponding to a prescribed algorithm. After performing these operations, new coded words are produced at the output of the digital filter, these new words corresponding to the filtered signal.

In a digital-to-analog converter (*DAC*) each coded word actuates a group of electronic switches which control the addition of reference voltages corresponding to each of the digits. As a result, the samples are reproduced in analog form at the *DAC* output. Such a decoding is a process reciprocal to the one occurring in the *ADC*.

The *DAC* output voltage $s'_{T,\ out}(t)$ has a stepped shape, the height of each step being equal to the sample of the output signal at the corresponding instant of time. Thus, by the discrete output signal $s_{T,\ out}(t)$ one should imply a sequence of "slender" pulses representing the discrete samples of the output signal $s_{out}(t)$.

Finally, in a two-port network, which could be called a synthesizing filter (*SF*), the discrete sequence is converted into a continuous output signal $s_{out}(t)$.

It is clear that all the above conversions that each sample undergoes must be performed during a time shorter than the interval T. In addition, it is necessary to provide for a strict synchronization of the operation of the electronic switches used for digit-by-digit addition, subtraction and other operations with coded words. All this requires a complex system of synchronization of auxiliary pulse trains used for erasing old information in the binary elements (e.g., flip-flops) and inserting new information during each interval T.

This problem is solved by shaping these trains from a single harmonic oscillation with a frequency $1/T$ produced by a reference oscillator. Since T is the main parameter of a digital filter, special attention is given to the frequency stability of the reference oscillator.

These complex problems are being successfully solved by means of modern integrated microcircuits.

Digital filters have important advantages. Their main advantages — operational reliability and the stability of parameters that cannot be obtained in analog filters — are due to the conversion of

a continuous signal into a binary number represented by standard signals (pulses and pauses).

Some other important advantages will be noted later in the text, after discussing the main characteristics of a digital filter in more detail.

It should be noted that in the analysis of the principle of operation of the circuit shown in Fig. 13.1, the digital-to-analog and analog-to-digital conversion are of no principal importance. One may assume that analog samples are inserted into a computer where they are subjected to mathematical processing (there are discrete systems of the analog type, in which digital coding is not used). In this connection, the next sections of this chapter deal with the operation of discrete systems first without taking into account *ADC* and *DAC*. The estimation of the errors associated with the quantization of the samples and also some other specific features of digital processing of signals are discussed in Sec. 13.11 through 13.13.

13.2. ALGORITHM OF DISCRETE CONVOLUTION (IN THE TIME DOMAIN)

A system for discrete processing of continuous signals (without digital coding) is shown in Fig. 13.2.

The process of sampling the input signal is conveniently regarded as the multiplication of a function $s(t)$ by a periodic train $y_T(t)$

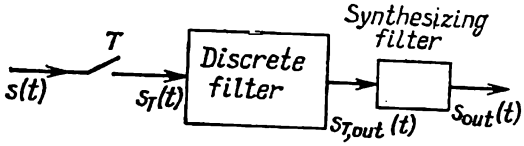


Fig. 13.2. Circuit for discrete processing of signals

of clock pulses. Such pulses are usually represented by delta functions, so that the function $y_T(t)$ can be defined by the expression

$$y_T(t) = \sum_{k=-\infty}^{\infty} \delta(t-kT) \tag{13.1}$$

The graph of this function is shown in Fig. 13.3*b* (the numeral one stands for the area of a clock pulse).

In this case the sampled signal

$$s_T(t) = s(t)\, y_T(t) = s(t) \sum_{k=-\infty}^{\infty} \delta(t-kT)$$

$$= \sum_{k=-\infty}^{\infty} s(kT)\, \delta(t-kT) = \sum_{k=-\infty}^{\infty} s_k \tag{13.2}$$

where

$$s_k = s(kT)\delta(t - kT) \tag{13.3}$$

With such a designation, s_k has the meaning of an infinitely short pulse located on the time axis at the point $t = kT$ and having an area numerically equal to the sample $s(kT)$. Thus, the samples

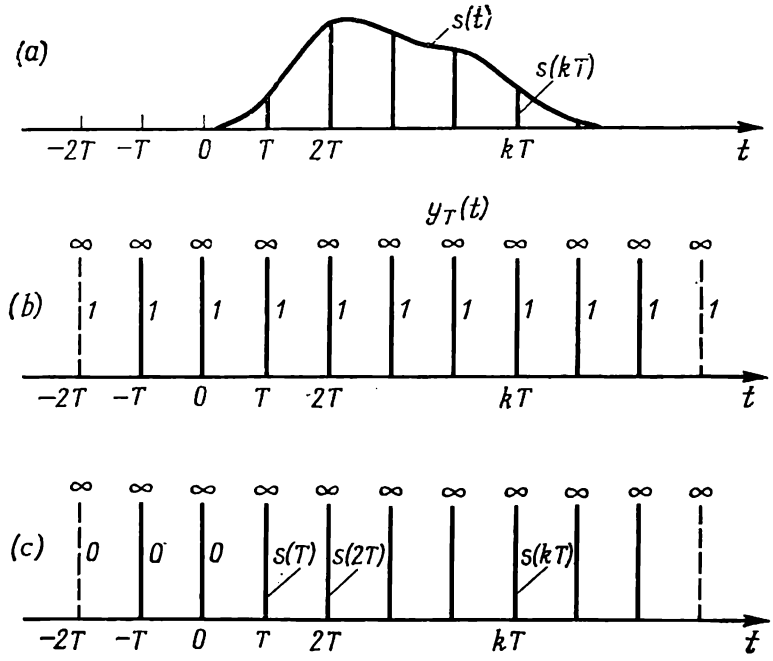


Fig. 13.3. Signal sampling

(a) original continuous signal; (b) train of clock pulses; (c) discrete samples of the signal

$s(kT)$ of the signal $s(t)$ (Fig. 13.3a) are the weighting coefficients of the delta functions (Fig. 13.3c)*.

The sampled signal $s_T(t)$ is fed to a discrete filter whose effect on $s_T(t)$ must be equivalent to the effect of some analog filter on $s(t)$. Let the impulse response $g(t)$ of the latter filter be known. Then the signal at the output of the analog filter is determined by the convolution [see expression (6.12)]

$$s_{out}(t) = \int_{-\infty}^{\infty} s(t-x)\, g(x)\, dx \tag{13.4}$$

* It should be borne in mind that the dimension of $s_T(t)$ differs from that of $s(t)$ since the function $\delta(t)$ has the dimension $1/t$. To maintain the dimension, expression (13.3) should be written in the form $s_k = T_s(kT)\,\delta(t - kT)$, which is equivalent to equating the area of the pulse s_k to the area of a square pulse with a base T and amplitude $s(kT)$ [4].

The interval T being appropriately selected, the discrete equivalent of the convolution can be written in the form

$$s_{out}(nT) = \sum_{k=-\infty}^{\infty} s[(n-k)T]\, g(kT) \qquad (13.5)$$

where $s_{out}(nT)$ is the nth sample of $s_{out}(t)$, the pulse corresponding to this sample, in accord with (13.3), being

$$s_{out,\,n} = s_{out}(nT)\,\delta(t-nT) \qquad (13.6)$$

In expression (13.5), $g(kT)$ has the meaning of the kth sample of the impulse response of the analog filter shown in Fig. 13.4 by

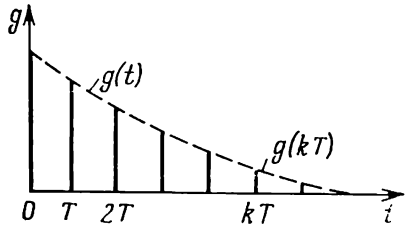


Fig. 13.4. Sampling the impulse response curve of an analog filter

a dashed line.

Expression (13.5) can be regarded as the algorithm of a discrete processing that is equivalent to the transmission of the signal through an analog filter with a given impulse response $g(t)$. If time is measured from the beginning of the input signal, expression (13.5) assumes the form

$$s_{out}(nT) = \sum_{k=0}^{n} s[(n-k)T] \times g(kT) \qquad (13.5')$$

The upper summation limit stems from the condition that $s[(n-k)T] = 0$ for $k > n$ and the lower limit, from the condition that $g(t) = 0$ for $t < 0$.

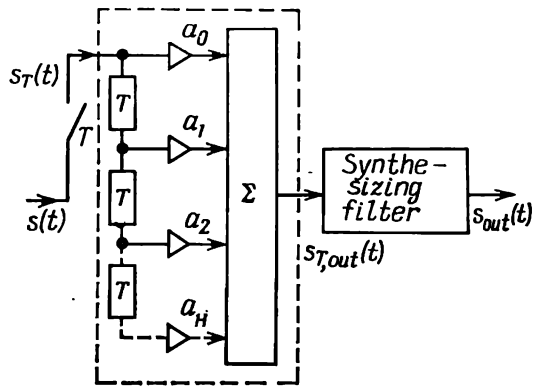


Fig. 13.5. Discrete filter

A schematic diagram of the device having the required discrete impulse response $g_T(t)$ is shown in Fig. 13.5. In this diagram, the letters a_0, a_1, a_2, . . ., a_H denote frequency-independent amplification factors and T is an ideal delay line, the delay coinciding with the rate of sampling of the signal $s(t)$. By appropriately selecting the

factors $a_0 = g(0)$, $a_1 = g(T)$, $a_2 = g(2T)$, . . . one can in principle realize very complex characteristics $g_T(t)$.

It is clear that the impulse response of the system under consideration should be written in the form

$$g_T(t) = g(0)\,\delta(t) + g(T)\,\delta(t-T) + g(2T)\,\delta(t-2T) + \ldots = a_0\delta(t) + a_1\delta(t-T) + a_2\delta(t-2T) + \ldots \quad (13.7)$$

Thus, $a_k = g(kT)$ is merely the weighting coefficient multiplying the delta function $\delta(t-kT)$. When a discrete signal $s_T(t)$ is applied to the input of the circuit enclosed in a dashed box in Fig. 13.5, a discrete pulse train $s_{T,out}(t)$ is produced at the output of the adder Σ, and the filtered continuous signal $s_{out}(t)$ is produced at the output of the synthesizing filter.

The realization of a discrete filter directly according to the diagram shown in Fig. 13.5 is possible only for comparatively simple signals having a small base $N = 2f_mT_s$ ($T_s = NT$ is the duration of the signal realization being processed). However, algorithm (13.5) is widely used for modelling filters by means of a computer.

The basic difficulty in the realization of a discrete filter consists in the realization of the memory element T. This difficulty is eliminated when using binary elements (flip-flops) for storing the signal for any required interval of time. The memory T is an important parameter of a discrete filter. This parameter has a decisive effect on the main filter characteristics — the transfer function and the impulse response.

13.3. DISCRETE FOURIER TRANSFORMS

To understand the specific features of discrete filtering, it is important to find out the structure of the spectrum of the sampled signal $s_T(t)$.

Let the spectral density $S(\omega)$ of the original (continuous) signal $s(t)$ and the sampling interval T be known. Regarding $s_T(t)$ as the product $s(t)y_T(t)$, where $y_T(t)$ is defined by expression (13.1), we find the spectral density of the function $s_T(t)$ by the formula

$$S_T(\omega) = \int_{-\infty}^{\infty} s(t)\,y_T(t)\,e^{-i\omega t}\,dt \quad (13.8)$$

A periodic train of delta-pulses can be represented in the form of a Fourier series

$$y_T(t) = \frac{1}{T}\sum_{n=-\infty}^{\infty} e^{in\omega_1 t},\quad \omega_1 = 2\pi/T \quad (13.9)$$

The coefficients of this series are equal to $1/T$, because the spectral density of a single delta-pulse is equal to unity, while the repetition interval of the pulses is equal to T [see formula (2.55)].

Substituting (13.9) into (13.8), we obtain

$$\mathbf{S}_T(\omega) = \frac{1}{T}\int_{-\infty}^{\infty} s(t) \sum_{n=-\infty}^{\infty} e^{in\omega_1 t} e^{-i\omega t}\, dt$$

$$= \frac{1}{T}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} s(t)\, e^{-i(\omega - n\omega_1)t}\, dt = \frac{1}{T}\sum_{n=-\infty}^{\infty} \mathbf{S}\left(\omega - n\frac{2\pi}{T}\right) \quad (13.10)$$

Thus, the spectrum $\mathbf{S}_T(\omega)$ of the *sampled signal is a series of successive spectra** $\mathbf{S}(\omega)$ *of the original signal* $s(t)$ *shifted by an amount* $2\pi/T$ *with respect to one another.*

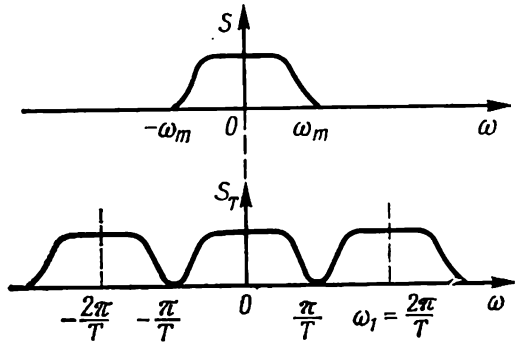


Fig. 13.6. Spectrum of a sampled signal

If the sampling interval satisfies the condition $T < 1/2f_m$ (in accord with the sampling theorem) and consequently, $\omega_m \leqslant \pi/T$, the individual spectra do not overlap, as shown in Fig. 13.6, and can be separated by means of filters at the system output.

The spectrum of the sampled signal assumes a periodic form. Expression (13.10) is useful for finding the relation between $\mathbf{S}(\omega)$ and $\mathbf{S}_T(\omega)$, however, in the general case where the relation between T and $\mathbf{S}(\omega)$ is arbitrary and the spectra may overlap, the use of formula (13.10) becomes difficult. Furthermore, it is desirable to have a formula allowing one to find $\mathbf{S}_T(\omega)$ directly by the specified time samples $s(kT)$ without resorting to the spectrum $\mathbf{S}(\omega)$ of the original continuous signal. Such a formula can easily be obtained by applying the Fourier transformation to expression (13.2). With the time being measured from the first sample $s(0)$, we have

$$\mathbf{S}_T(\omega) = \int_0^{\infty} s_T(t)\, e^{-i\omega t}\, dt = \int_0^{\infty} e^{-i\omega t} \sum_{k=0}^{\infty} s(kT)\, \delta(t - kT)\, dt$$

$$= \sum_{k=0}^{\infty} s(kT) \int_0^{\infty} e^{-i\omega t}\delta(t - kT)\, dt = \sum_{k=0}^{\infty} s(kT)\, e^{-i\omega kT} \quad (13.11)$$

* The quantity $\mathbf{S}_T(\omega)$ differs in dimension from $\mathbf{S}(\omega)$, as well as the originals $s_T(t)$ and $s(t)$ (see the footnote on p. 503).

A real signal is approximated by means of a finite number of samples. With this number being equal to N, expression (13.11) assumes the form

$$S_T(\omega) = \sum_{k=0}^{N-1} s(kT)\, e^{-i\omega kT} \tag{13.12}$$

The use of a computer requires that the signal be sampled in both the time and the frequency domain. In the latter case, the frequency spectrum $S_T(\omega)$ is defined by a set of its samples $S_T(n\omega_1)$ taken at discrete frequencies $\omega = n\omega_1$.

In Sec. 2.15 it was found that the number of degrees of freedom of the signal is the same both in time and in frequency. The frequency interval ω_1 between adjacent samples must be taken at $\omega_1 = 2\omega_m/N$. Since $\omega_m = \pi/T$, then $\omega_1 = 2\pi/NT$. This relation agrees with the definition $\Delta\omega = 2\pi/T_s$ given in Sec. 2.15, because the product NT has the meaning of the duration T_s of the original continuous signal.

Substituting $\omega = n\omega_1$ into expression (13.12), we obtain a formula for determining *frequency samples*:

$$S_T(n\omega_1) = \sum_{k=0}^{N-1} s(kT)\, e^{-i\omega_1 nkT} = \sum_{k=0}^{N-1} s(kT)\, e^{-i\frac{2\pi}{N}nk}$$

$$n = 0, \pm 1, \pm 2, \ldots, \pm (N-1)/2 \tag{13.13}$$

Relation (13.13) is known as the *discrete Fourier transform*.

As $|n|$ grows in excess of $(N-1)/2$, the function $S_T(n\omega_1)$ recurs periodically. Therefore, $S_T(-\omega_1)$ can be equated to $S_T[(N-1)\omega_1]$, $S_T(-2\omega_1) = S_T[(N-2)\omega_1]$, respectively, etc. This makes it possible to write expression (13.13) in a somewhat modified form convenient for computer calculations:

$$S_T(n\omega_1) = \sum_{k=0}^{N-1} s(kT)\, e^{-i\frac{2\pi}{N}nk},$$

$$n = 0, 1, 2, \ldots, N-1 \tag{13.13'}$$

The numbering of the samples for odd and even values of N is explained in Figs. 13.7 (for $N = 7$) and 13.8 (for $N = 8$). In these examples, the function $s(t)$ is supposed to be real, therefore, the frequency samples located symmetrically with respect to the point $n = 0$ ($\omega = 0$) must form complex-conjugate pairs. In order to satisfy this condition the samples $S_T(n\omega_1)$ must be arranged in the middle of the respective frequency intervals (shaded areas in Fig. 13.7 and Fig. 13.8). The lower parts of these figures show the serial numbers (bracketed) of the samples, corresponding to the negative values of n after shifting the samples to the right for N frequency intervals. The total width of the spectrum is $2\omega_m = N\omega_1$ both for even and for odd N.

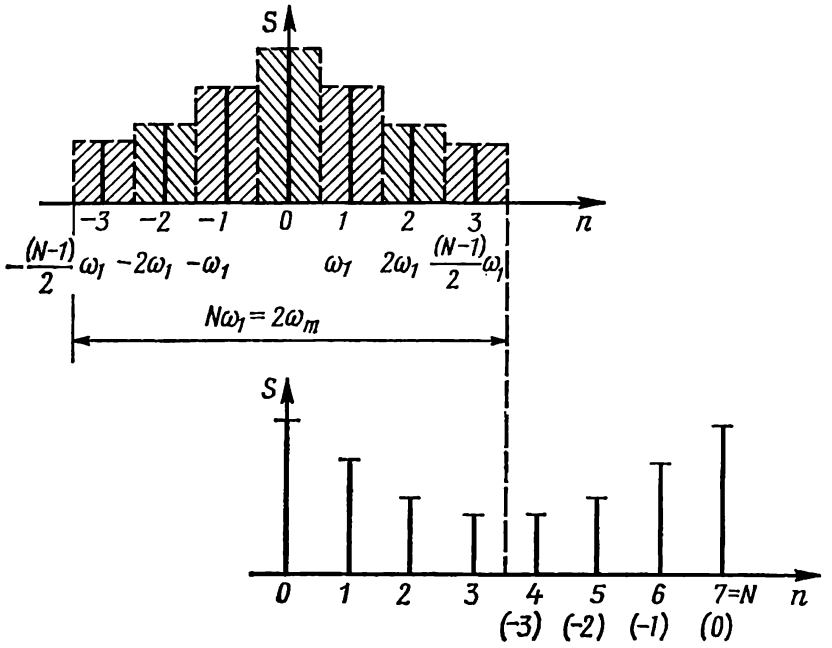


Fig. 13.7. Discrete Fourier transform. Numbering of samples for odd N

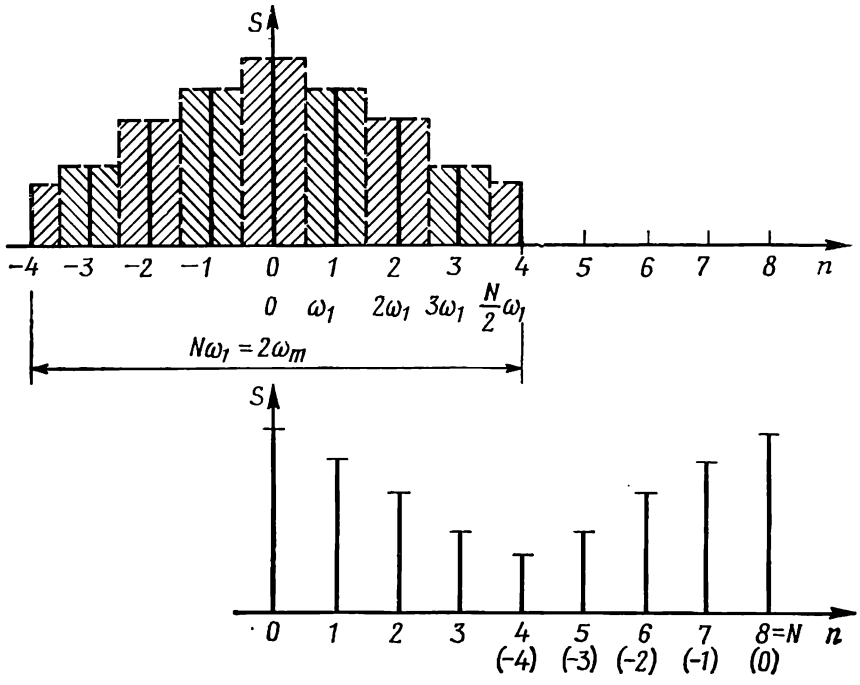


Fig. 13.8. Numbering of samples for even N

It is also possible to introduce the concept of a discrete inverse Fourier transform. By analogy with the pair of Fourier transforms (2.48) and (2.49), the discrete inverse transform may be defined as

$$s(kT) = C \sum_{n=0}^{N-1} \mathbf{S}_T(n\omega_1)\, e^{i\frac{2\pi}{N}nk}, \quad k = 0, 1, 2, \ldots, N-1$$

To find the constant coefficient C, let us substitute $\mathbf{S}_T(n\omega_1)$ from (13.13′) into the above expression:

$$s(kT) = C \sum_{n=0}^{N-1} \left(\sum_{m=0}^{N-1} s(mT)\, e^{-i\frac{2\pi}{N}nm} \right) e^{i\frac{2\pi}{N}nk}$$

$$= C \sum_{m=0}^{N-1} s(mT) \sum_{n=0}^{N-1} e^{i\frac{2\pi}{N}n(k-m)}$$

For $m = k$, the internal sum turns into N, and for any other value of m, it vanishes (as the sum of vectors whose ends divide the circle of a unit radius into equal arcs). Therefore, we have only one term $Cs(kT)\,N$ on the right-hand side so that we have the equality $C = 1/N$. Thus, the discrete inverse Fourier transform assumes the following form:

$$s(kT) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{S}_T(n\omega_1)\, e^{i\frac{2\pi}{N}kn}, \quad k = 0, 1, 2, \ldots, N-1 \quad (13.14)$$

As in the case of the discrete direct Fourier transform, outside the interval $0 \leqslant k \leqslant N-1$ the function $s(kT)$ recurs periodically.

Some properties of continuous Fourier transforms discussed in Sec. 2.7 can also be readily worded for a discrete Fourier transform.

1. Linearity of the transform: the spectrum of the sum (difference) of discrete signals is equal to the sum (difference) of their spectra.

2. Shift of a discrete signal in time. Repeating the arguments leading to expression (2.57), one can easily show that if a signal $s(t)$, represented by a set of samples $s(kT)$, $k = 0, 1, 2, \ldots, N-1$, corresponds to a discrete Fourier transform $\mathbf{S}_T(n\omega_1)$, a signal $s(t+mT)$, where m is an integer, corresponds to a discrete Fourier transform $e^{i\frac{2\pi}{N}nm}\mathbf{S}_T(n\omega_1)$. In other words, a shift in the sequence of samples for m intervals only changes the phase-frequency characteristic of the discrete Fourier transform by an amount $\frac{2\pi}{N}nm$ (the delay theorem).

3. Convolution theorem. If a discrete Fourier transform $\mathbf{S}_T(n\omega_1)$ corresponds to a discrete signal $s_T(t) = \sum_{k=0}^{N-1} s(kT)\delta(t-kT)$ and a discrete Fourier transform $\mathbf{G}_T(n\omega_1)$ corresponds to a signal $g_T(t) = \sum_{k=0}^{N-1} g(kT)\,\delta(t-kT)$, the product $\mathbf{S}_T(n\omega_1)\mathbf{G}_T(n\omega_1)$

corresponds to the signal

$$y(mT) = \sum_{k=0}^{N-1} s[(m-k)T]\, g(kT)$$

The derivation of this expression is similar to the derivation of (2.64) [see also (13.7)].

In the preceding chapters it was noted that the duration and the spectrum width of a signal cannot be limited simultaneously. Representation of a signal by a finite number N of pulses and a finite number N of frequency samples inevitably results in some distortion of the shape of the signal. However, this distortion appears not during the transformation from $s(kT)$ to $S_T(n\omega_1)$ or during the inverse transformation from $S_T(n\omega_1)$ to $s(kT)$ by means of transforms (13.13) and (13.14), but during the transformation from the discrete to the continuous representation. This question is discussed in the Ses. 13.4.

13.4. ERROR OF SAMPLING SIGNALS OF FINITE LENGTH

In Sec. 2.14 it was noted that the signal cannot be limited in duration and spectrum width simultaneously. This is especially manifest in sampling relatively short pulse signals.

Let us specify a signal $s(t)$ of duration T_s and, strictly speaking, of infinitely wide spectrum $S(\omega)$. When selecting the sampling interval T by the Kotelnikov theorem (see Sec. 2.14), an uncertainty arises in the estimation of ω_m — the boundary frequency of the signal spectrum. The selection of this frequency determines the sampling interval $T = \pi/\omega_m$, and with a given signal duration T_s, the number of the discrete samples is $N = T_s/T = T_s\omega_m/\pi$.

To restore the continuous signal, one requires a filter with a passband $|\omega| = \omega_m$ (see Sec. 13.3). To simplify the analysis, we proceed from an ideal filter with an amplitude-frequency characteristic

$$K(i\omega) = \begin{cases} 1 \text{ for } |\omega| < \omega_m \\ 0 \text{ for } |\omega| \geqslant \omega_m \end{cases}$$

The phase-frequency characteristic of such a filter is linear within the passband and only delays the signal, and this delay will not be taken into account in further analysis.

The filter with such a passband rejects those components of the spectrum $S(\omega)$, whose frequencies ω exceed ω_m. Therefore, by the sampling error one can imply the time function corresponding to the spectrum portion being cut off. Let this function be denoted $\Delta s(t)$. Then

$$\Delta s(t) = \frac{1}{2\pi}\int_{-\infty}^{-\omega_m} S(\omega)\, e^{i\omega t}\, d\omega + \frac{1}{2\pi}\int_{\omega_m}^{\infty} S(\omega)\, e^{i\omega t}\, d\omega$$

Let us illustrate this expression by an example of a simple signal in the form of a square pulse (Fig. 13.9*a*).

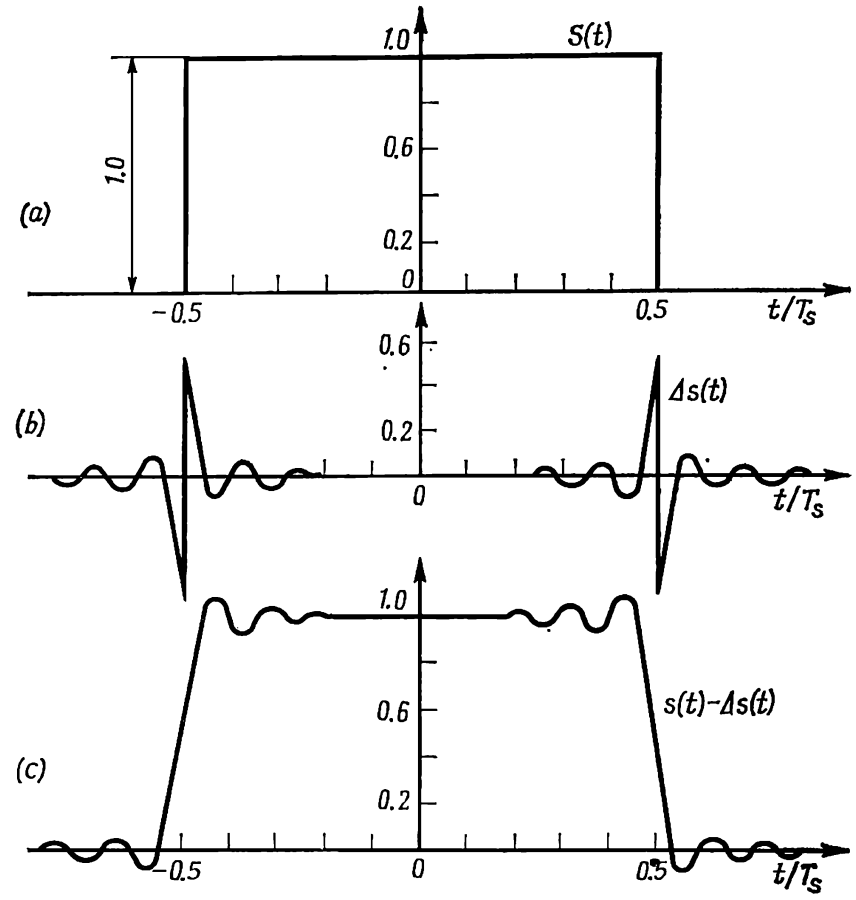


Fig. 13.9. Distortion of the shape of a pulse due to its spectrum being limited to the boundary frequency $2f_m T_s = 16$

(a) original pulse; (b) error voltage; (c) distorted pulse

The spectral density of this signal $s(t)$, equal to [see (2.68)]

$$S(\omega) = AT_s \frac{\sin(\omega T_s/2)}{\omega T_s/2}$$

is an even function of ω. Therefore, the expression for $\Delta s(t)$ can easily be reduced to the form

$$\Delta s(t) = \frac{A}{\pi}\left[\int_0^{\infty} \frac{1}{\omega} \sin\omega\left(t+\frac{T_s}{2}\right) d\omega - \int_0^{\omega_m} \frac{1}{\omega} \sin\omega\left(t+\frac{T_s}{2}\right) d\omega - \int_0^{\infty} \frac{1}{\omega} \sin\omega\left(t-\frac{T_s}{2}\right) d\omega + \int_0^{\omega_m} \frac{1}{\omega} \sin\omega\left(t-\frac{T_s}{2}\right) d\omega\right]$$

Taking into account the well-known relations

$$\int_0^{\infty} \frac{\sin ax}{x}\,dx = \begin{cases} \pi/2 \text{ for } a > 0 \\ -\pi/2 \text{ for } a < 0 \end{cases}$$

we come to the following result:

$$\Delta s(t) = \begin{cases} \frac{A}{\pi}\left[\pi - \mathrm{Si}\,\frac{\omega_m T_s}{2}\left(1+\frac{2t}{T_s}\right) - \mathrm{Si}\,\frac{\omega_m T_s}{2}\left(1-\frac{2t}{T_s}\right)\right] \text{for } |t| < \frac{T_s}{2} \\ \frac{A}{\pi}\left[-\mathrm{Si}\,\frac{\omega_m T_s}{2}\left(1+\frac{2t}{T_s}\right) - \mathrm{Si}\,\frac{\omega_m T_s}{2}\left(1-\frac{2t}{T_s}\right)\right] \text{ for } |t| \geqslant \frac{T_s}{2} \end{cases}$$

In these expressions, $\mathrm{Si}\,y = \int_0^y \frac{\sin ax}{x}\,dx$ is the sine integral. Let N samples be allotted to sample a pulse. In this case, $T = T_s/N$, $\omega_m = \pi/T = \pi N/T_s$, $\omega_m T_s/2 = N\pi/2$, and the preceding expression is written in the form

$$\Delta s(t) = \begin{cases} A\left\{1 - \frac{1}{\pi}\mathrm{Si}\left[N\frac{\pi}{2}\left(1+\frac{2t}{T_s}\right)\right] - \frac{1}{\pi}\mathrm{Si}\left[N\frac{\pi}{2}\left(1-\frac{2t}{T_s}\right)\right]\right\} \text{ for } |t| < \frac{T_s}{2} \\ A\left\{-\frac{1}{\pi}\mathrm{Si}\left[N\frac{\pi}{2}\left(1+\frac{2t}{T_s}\right)\right] + \frac{1}{\pi}\mathrm{Si}\left[N\frac{\pi}{2}\left(\frac{2t}{T_s}-1\right)\right]\right\} \text{ for } |t| \geqslant \frac{T_s}{2} \end{cases}$$

The graphs of the function $\Delta s\,(t)$ and of the signal at the filter output for $A = 1$ are shown in Fig. 13.9*b* and *c*. It is interesting to note that for a square pulse, the character of the fuction $\Delta s\,(t)$ is determined solely by the parameter $N = 2f_m T_s$.

The function $\Delta s\,(t)$ can also be defined by analyzing the action on a synthesizing filter of a train of physical pulses correspoding to the samples of the signal $s\,(t)$. The response of the filter to a sample $s\,(nT)$ is equal to $s\,(nT)\,g\,(t - nT)$, where $g\,(t)$ is the impulse response which, for $\mathbf{K}\,(i\omega) = 1$, $|\,\omega\,| \leqslant \omega_m$, is defined by the

expression

$$g(t) = \frac{1}{2\pi} \int_{-\omega_m}^{\omega_m} e^{i\omega t}\, d\omega = \frac{1}{\pi} \frac{\sin \omega_m t}{t} = \frac{1}{T} \frac{\sin \omega_m t}{\omega_m t} = \frac{1}{T} \operatorname{sinc}(\omega_m t)$$

The function sinc $(\omega_m t)$ is the basic function of the Kotelnikov series (see Sec. 2.15). Therefore, the function $\Delta s(t)$ existing after the signal $s(t)$ has ceased can be found by adding together the functions $s(nT)g(t - nT)$ of all the samples preceding the instant $t = T_s/2$. At the discrete points $t = kT$ the function $\Delta s(t)$ is equal to zero. From this it follows that an increase in the intervals T results in distortion of the signal only between the sampling points. It is easy to show that the duration of the function $\Delta s(t)$ practically amounts to 3-4 cycles. Therefore, an increase in the number N of the samples reduces this fraction of time of the total pulse duration.

13.5. DISCRETE LAPLACE TRANSFORMS

Transforming to a new variable $p = \sigma + i\omega$ in expression (13.11), one can, as in Sec. 2.13, introduce the concept of a *discrete Laplace transform*:

$$\mathbf{S}_T(p) = \sum_{k=0}^{\infty} s(kT)\, e^{-pkT} \tag{13.15}$$

The determination of the original, i.e., the function $s_T(t)$, by the given transform $\mathbf{S}_T(p)$ can be effected by means of a *discrete inverse Laplace transform*, which is written in the common form

$$s(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \mathbf{S}_T(p)\, e^{pt}\, dp \tag{13.16}$$

coinciding with (2.107).

Let us demonstrate that the substitution of $\mathbf{S}_T(p)$ from formula (13.15) into (13.16) gives the required result:

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \sum_{k=0}^{\infty} s(kT)\, e^{-pkT} e^{pt}\, dp = \sum_{k=0}^{\infty} s(kT) \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{p(t-kT)}\, dp$$

Taking into account (see also Table 2.1) that as $c \to 0$

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{pt}\, dp \to \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t}\, d\omega = \delta(t) \tag{13.17}$$

the right-hand side of the preceding expression can easily be reduced to the form

$$\sum_{k=0}^{\infty} s(kT)\, \delta(t - kT) = s_T(t) \tag{13.18}$$

coinciding with (13.2). Thus, we make certain that expression (13.16) defines the entire sequence of samples of the signal $s(t)$.

To find a single, nth sample $s(nT)$, without the factor $\delta(t - nT)$, one can use a simpler expression

$$s(nT) = T\frac{1}{2\pi i}\int\limits_{c-i\pi/T}^{c+i\pi/T} \mathbf{S}_T(p)\, e^{pnT}\, dp \qquad (13.19)$$

in which integration is performed over a single frequency interval from $-\pi/T$ to π/T.

Indeed, substituting, as before, $\mathbf{S}_T(p)$ by formula (13.15) into (13.19), we obtain

$$s(nT) = T\frac{1}{2\pi i}\int\limits_{c-i\pi/T}^{c+i\pi/T}\left(\sum_{k=0}^{\infty} s(kT)\, e^{-pkT}\right) e^{pnT}\, dp$$

$$= T\sum_{k=0}^{\infty} s(kT)\frac{1}{2\pi i}\int\limits_{-i\pi/T}^{c+i\pi/T} e^{p(n-k)T}\, dp \qquad (13.20)$$

Let us compute the integral

$$\frac{1}{2\pi i}\int\limits_{c-i\pi/T}^{c+i\pi/T} e^{p(n-k)T}\, dp = e^{c(n-k)T}\frac{1}{2\pi}\int\limits_{-\pi/T}^{\pi/T} e^{i\omega(n-k)T}\, d\omega$$

$$= e^{c(n-k)T}\frac{1}{\pi}\frac{\sin(n-k)\pi}{(n-k)T} \qquad (13.21)$$

For $n = k$ this expression is equal to $1/T$ and for $n \neq k$ it is equal to zero. Therefore, only one term, $s(nT) = s(kT)$, remains on the right-hand side of (13.20).

Where the discrete series $s_T(t)$ is nonzero for negative indices k, one has to use a *discrete bilateral Laplace transform.*

By analogy with expressions (2.108), (2.112), and (2.114), let us represent this transform in the form

$$\mathbf{S}_T'(p) = \sum_{k=0}^{\infty} s(kT)\, e^{-pkT} + \sum_{k=-\infty}^{0} s(kT)\, e^{-pkT} - s(0) \qquad (13.22)$$

where $s(0)$ compensates for the index $k = 0$ in both sums.

Setting $k = -m$ in the second sum, we obtain

$$\mathbf{S}_T(p) = \sum_{k=0}^{\infty} s(kT)\, e^{-pkT} + \sum_{m=0}^{\infty} s(-mT)\, e^{pmT} - s(0)$$

$$= \mathbf{S}_{T+}(p) + \mathbf{S}_{T-}(p) - s(0) \qquad (13.23)$$

In the particular case of even function $s(t)$, where $s(-mT) = s(mT)$, relation similar to (2.116) holds:

$$\mathbf{S}_{T-}(p) = \mathbf{S}_{T+}(-p)$$

Then,

$$S_T(p) = S_{T+}(p) + S_{T+}(-p) - s(0) = \sum_{k=-\infty}^{\infty} s(kT)\, e^{-pkT} \qquad (13.24)$$

In expressions (13.23) and (13.24), S_{T+} is the unilateral transform defined by expression (13.15).

It is not difficult to define the convergence region of series (13.23). This region is a band limited by quantities σ_1 and σ_2 (see Sec. 2.13, Fig. 2.33) which are determined by the properties of the continuous signal from which the discrete trains for $t > 0$ and $t < 0$ have been obtained.

13.6. TRANSFER FUNCTION OF A DISCRETE FILTER

Let us return to the diagram in Fig. 13.2 and set up an expression for the transfer function of a discrete filter in the form of the ratio

$$K_T(p) = S_{T,\,out}(p)/S_T(p) \qquad (13.25)$$

Here, $S_T(p)$ and $S_{T,\,out}(p)$ are the Laplace transforms respectively of the discrete input and output oscillations discussed in the preceding section.

The first of these functions is determined by unilateral Laplace transform (13.15). In order to set up a similar expression for $S_{T,out}(p)$, it is necessary to specify the algorithm of operation of the computer. First, let us consider a simple filter which does not use feedback. In accordance with expressions (13.5′) and (13.6) and the diagram in Fig. 13.5, the output pulse $s_{out,\,n}$ at the instant $t = nT$ is

$$s_{out,\,n} = \sum_{k=0}^{H} a_k s[(n-k)T]\,\delta(t-nT) \qquad (13.26)$$

where H is the number of stored and actually added preceding pulses of the input train (i.e., the number of the memory elements T in Fig. 13.5).

The complete train of output pulses is given by

$$s_{T,\,out}(t) = \sum_{n=0}^{\infty} s_{out,\,n} = a_0 s(0)\,\delta(t) + [a_0 s(T) + a_1 s(0)]\,\delta(t-T)$$

$$+ [a_0 s(2T) + a_1 s(T) + a_2 s(0)]\,\delta(t-2T) + \ldots$$

$$= a_0 \sum_{n=0}^{\infty} s(nT)\,\delta(t-nT)$$

$$+a_1\sum_{n=1}^{\infty} s[(n-1)T]\,\delta(t-nT)+\ldots$$

$$+a_H\sum_{n=H}^{\infty} s[(n-H)T]\,\delta(t-nT) \qquad (13.27)$$

The first sum on the right-hand side of this expression (with the coefficient a_0) is the input pulse train (13.2), the second sum is the same train delayed for the time T, the third one, that delayed for $2T$, etc. Consequently, the Laplace transform of expression (13.27) will be

$$\mathbf{S}_{T,\,out}(p)=a_0\mathbf{S}_T(p)+a_1\mathbf{S}_T(p)\,e^{-pT}+a_2\mathbf{S}_T(p)\,e^{-2pT}+\ldots$$

$$+a_n\mathbf{S}_T(p)\,e^{-HpT}=\mathbf{S}_T(p)\sum_{k=0}^{H} a_k e^{-kpT} \qquad (13.28)$$

By dividing expression (13.28) by $\mathbf{S}_T(p)$ and taking into account (13.25), we obtain

$$\mathbf{K}_T(p)=\sum_{k=0}^{H} a_k e^{-kpT} \qquad (13.29)$$

Setting $p = i\omega$, we find the transfer function as a function of frequency ω:

$$\mathbf{K}_T(i\omega)=\sum_{k=0}^{H} a_k e^{-ik\omega T} \qquad (13.30)$$

Expressions (13.29) and (13.30) also stem directly from the equivalent circuit shown in Fig. 13.5, it being assumed that the transmission coefficient of the adder Σ is equal to unity.

By properly selecting $a_0, a_1, \ldots, a_H$, one can synthesize filters with different amplitude-frequency and phase-frequency characteristics.

The periodicity of the function $\mathbf{K}_T(i\omega)$ can easily be demonstrated. In fact, by adding to the argument ω the quantity $n2\pi/T$ (where n is an integer) and taking into account that $e^{-ikn2\pi} = 1$, we obtain

$$\mathbf{K}_T[i(\omega+n2\pi/T)]=\sum_{k=0}^{H} a_k e^{-ik(\omega+n2\pi/T)T}$$

$$=\sum_{k=0}^{H} a_k e^{-ik\omega T}e^{-ikn2\pi}=\mathbf{K}_T(i\omega)$$

The transfer function of the discrete filter can be written in the

form*

$$\mathbf{K}_T(i\omega) = \sum_{n=-\infty}^{\infty} \mathbf{K}\left[i\left(\omega - n\frac{2\pi}{T}\right)\right] \tag{13.31}$$

where $\mathbf{K}(i\omega)$ is the transfer function of the same filter when the input signal is not sampled. Expression (13.31) is similar to (13.10).

Thus, the transfer function of a discrete filter has a periodic structure, like the spectra of the input signal $\mathbf{S}_T(i\omega)$ and the output signal

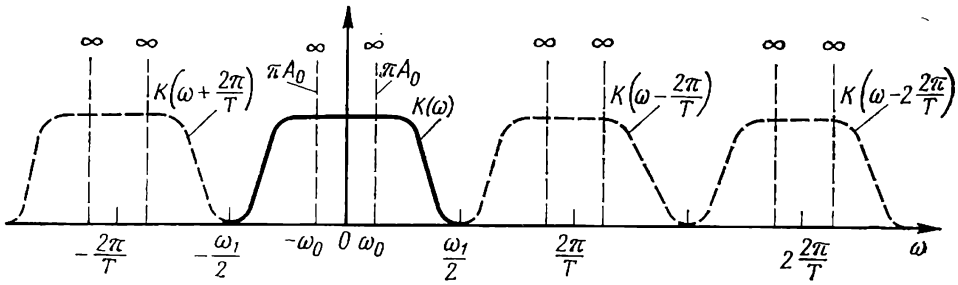


Fig. 13.10. Amplitude-frequency characteristic of a discrete filter

$\mathbf{S}_{T,\,out}(i\omega)$. This statement is illustrated in Fig. 13.10 for a low-pass filter whose input is fed with a harmonic oscillation $s(t) = A_0 \cos \omega_0 t$ with a spectral density $\pi A_0 [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]$.

The continuous line shows the amplitude-frequency characteristic of the filter in the central interval $(-\omega_1/2, \omega_1/2)$. After converting the discrete signal into a continuous signal (by means of the synthesizing filter shown in Fig. 13.1), it is exactly this frequency interval that defines the spectral composition of the output signal.

The transfer function $\mathbf{K}_T(p)$ can be used for finding the Laplace transform $\mathbf{S}_{T,\,out}(p) = \mathbf{S}_T(p)\,\mathbf{K}_T(p)$, and then the output (discrete) signal can be found by means of the inverse Laplace transform:

$$s_{T,\,out}(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \mathbf{S}_{T,\,out}(p)\, e^{pt}\, dp \tag{13.32}$$

Let us explain the use of the above expressions for a simple first-order filter shown in Fig. 13.11*a*. The impulse response of such a filter is a pair of pulses [see expression (13.7)]:

$$g_T(t) = a_0\delta(t) + a_1\delta(t - T) \tag{13.33}$$

* Expression (13.31) can be obtained by applying the Fourier transformation to the impulse response of the discrete filter, which can be written in the form

$$g_T(t) = T\sum_{k=0}^{\infty} g(kT)\,\delta(t - kT)$$

This expression differs from (13.7) by the factor T which is necessary for restoring the required dimension and for normalization (see the footnote on p. 503).

and the algorithm of the computer, according to expression (13.26), is given by

$$s_{out,\,n} = \sum_{k=0}^{n} a_k s\,[(n-k)\,T]\,\delta\,(t-nT) = \{a_0 s\,(nT) + a_1 s\,[(n-1)\,T]\}\,\delta\,(t-nT)$$

To this algorithm corresponds the transfer function

$$\mathbf{K}_T\,(p) = a_0 + a_1 e^{-pt} \tag{13.34}$$

[This result can also be obtained directly from expression (13.29).]

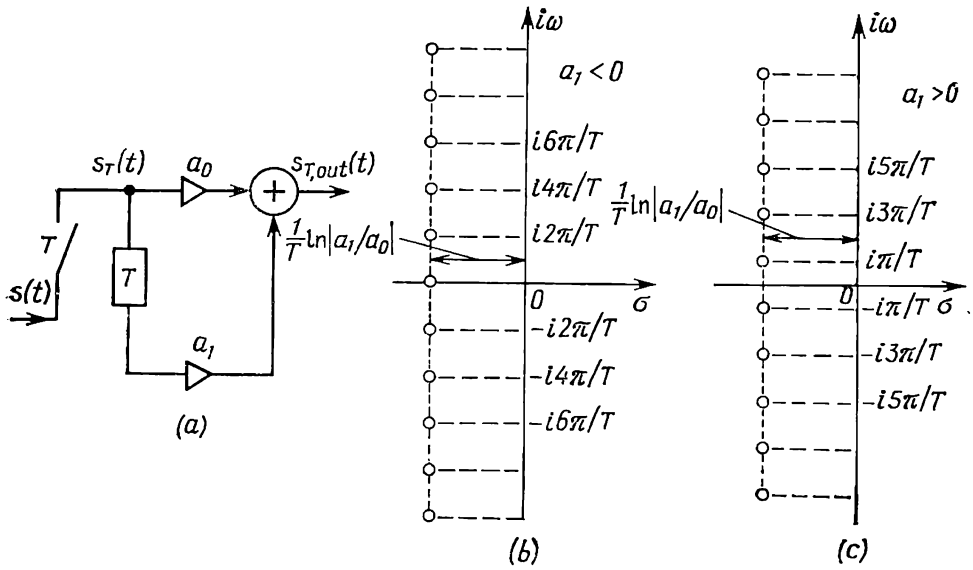


Fig. 13.11. (*a*) First-order discrete filter and the location of the zeros of its transfer function on the p-plane for (*b*) $a_1 < 0$ and (*c*) $a_1 > 0$

The zeros of the transfer function on the p-plane are determined as the roots of the equation $a_0 + a_1 e^{-pT} = 0$ or $e^{pT} = e^{\sigma T} e^{i\omega T} = -a_1/a_0$.

Denoting the roots by the expression $p_{0m} = \sigma_{0m} + i\omega_{0m}$, we obtain $e^{\sigma_{0m} T}\, e^{i\omega_{0m} T} = -a_1/a_0$ from which follow the equalities

$$e^{\sigma_{0m} T} = |-a_1/a_0|$$

$$\sigma_{0m} = \frac{1}{T} \ln |a_1/a_0|$$

$$\omega_{0m} T = \arg(-a_1/a_0) + im\,2\pi$$

where m is any integer.

The coefficients a_0 and a_1 are real numbers, it being assumed that $a_0 > 0$. Then, for $a_1 < 0$, $\arg(-a_1/a_0) = 0$ and $\omega_{0m} = m2\pi/T$. For $a_1 > 0$, $\arg(-a_1/a_0) = \pi$ and $\omega_{0m} = (2m+1)\pi/T$.

These values of ω_{0m}, together with σ_{0m}, determine the position of the zeros of the transfer function for $a_1 < 0$ and $a_1 > 0$ (Fig. 13.11).

From the expression

$$K_T(i\omega) = a_0 + a_1 e^{-i\omega T} \tag{13.35}$$

obtained by replacing p in (13.34) by $i\omega$, one can easily derive a formula for the amplitude-frequency characteristic

$$K_T(\omega) = \sqrt{a_0^2 + a_1^2 + 2a_0 a_1 \cos \omega T} = \sqrt{a_0^2 + a_1^2 + 2a_0 a_1 \cos\left(2\pi \frac{\omega}{\omega_1}\right)} \tag{13.36}$$

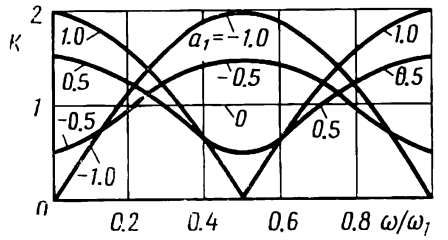


Fig. 13.12. Amplitude-frequency characteristic of a first-order digital filter (see Fig. 13.11*a*)

and for the phase-frequency characteristic

$$\varphi(\omega) = -\arctan \frac{a_1 \sin \omega T}{a_0 + a_1 \cos \omega T} = -\arctan \frac{a_1 \sin (2\pi\omega/\omega_1)}{a_0 + a_1 \cos (2\pi\omega/\omega_1)} \tag{13.37}$$

In these expressions, $\omega_1 = 2\pi/T = 2\pi f_1$ is the angular repetition frequency of the pulses when sampling the signal.

The amplitude-frequency characteristics for several values of a_1 are shown in Fig. 13.12. Outside the frequency interval 0, ω_1 the characteristics must be continued periodically. From Fig. 13.12 it is clear that when $a_1 = -1$ and $a_0 = 1$, the filter can be used for suppressing oscillations of frequencies $\omega = 0$, $\omega = \omega_1$, $\omega = 2\omega_1$, . . ., and when $a_1 = 1$, for suppressing those of frequencies $0.5\omega_1$, $1.5\omega_1$, $2.5\omega_1$ . . .

Such filters are often called *comb rejection filters*.

13.7. TRANSFER FUNCTION OF A RECOURSIVE FILTER

The capabilities of a filter are widened if feedback loops are introduced into the circuit shown in Fig. 13.5 (see Fig. 13.13). (Later in the text it will be shown how such a circuit can be simplified by using memory elements T both for feedforward and for feedback.) The level of the signal at the adder output at any instant nT depends not only on the H samples of the *input signal*, but also on some amount of the *output signal* samples taken at preceding instants. Such filters are called *recoursive*. For a recoursive filter, expression (13.26) should be replaced by a more general expression

$$s_{out,n} = \{a_0 s(nT) + a_1 s[(n-1)T] + a_2 s[(n-2)T] + \ldots + a_H s[(n-H)T]\}\,\delta(t-nT) + \{b_1 s_{out}[(n-1)T] +$$

$$+ b_2 s_{out}[(n-2)T] + \ldots$$
$$+ b_M s_{out}[(n-M)T]\} \delta(-nT) \qquad (13.38)$$

where M is the number of the preceding *output pulses* being added.

By analogy with (13.27) and (13.28), one can easily obtain the Laplace transform for the entire train of input pulses:

$$S_{T,out}(p) = S_T(p) \sum_{k=0}^{H} a_k e^{-kpT} + S_{T,out}(p) \sum_{k=1}^{M} b_k e^{-kpT} \qquad (13.39)$$

whence

$$K_T(p) = S_{T,out}(p)/S_T(p) = \sum_{k=0}^{H} a_k e^{-kpT} \Big/ \left(1 - \sum_{k=1}^{M} b_k e^{-kpT}\right)$$
$$= \alpha_T(p)/\beta_T(p) \qquad (13.40)$$

The function thus obtained can be treated as the transfer function of two cascade-connected filters: one with a transfer function $1/\beta_T(p)$

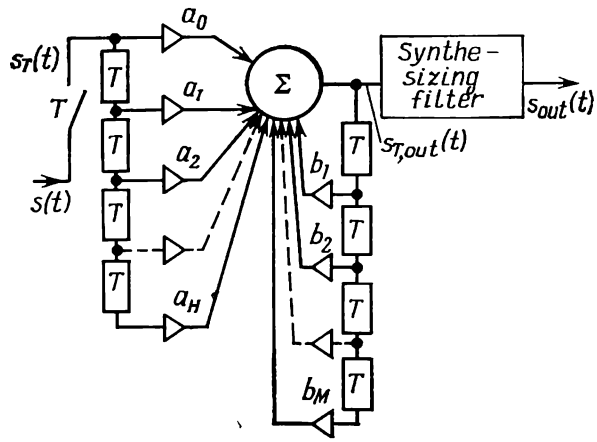


Fig. 13.13. Digital filter with feedback loops

and the other with a transfer function $\alpha_T(p)$. Such representation is illustrated by the canonical diagram shown in Fig. 13.14. The number of the memory elements T in this diagram is half that in the circuit of Fig. 13.13.

Recoursive filters make it possible to obtain frequency characteristics inherent in filters whose transfer functions on the plane $p = \sigma + i\omega$ have not only zeros (as that of the circuit shown in Fig. 13.2) but poles as well.

Let us illustrate expression (13.39) by an example of a simple filter in which only one preceding pulse is stored. The algorithm of such

a filter [see (13.38)] assumes the form

$$s_{out,n} = \{s(nT) + b_1 s_{out}[(n-1)T]\}\,\delta(t-nT) \qquad (13.41)$$

and its circuit diagram is shown in Fig. 13.15.

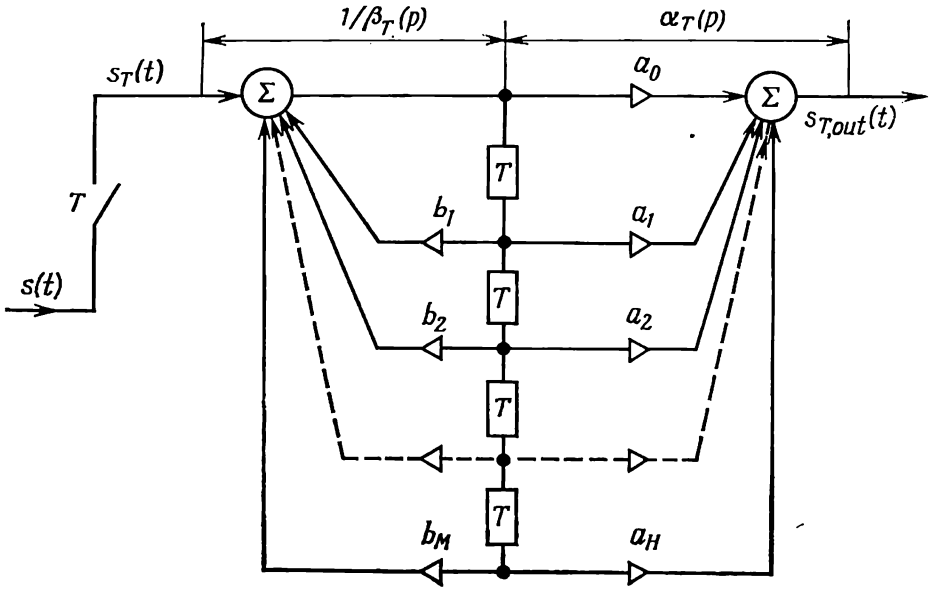


Fig. 13.14. Canonical diagram of a digital recoursive filter

According to formula (13.40), the transfer function of this filter is

$$\mathbf{K}_T(p) = 1/(1 - b_1 e^{-pT});\ \mathbf{K}_T(i\omega) = 1/(1 - b_1 e^{-i\omega T}) \qquad (13.42)$$

The poles of the transfer function are located at the points

$$p_m = \frac{1}{T}(\ln|b_1| + i2\pi m),\ m = 0,\ \pm 1,\ \pm 2,\ \pm \ldots$$

(Fig. 13.16*a*, *b*).

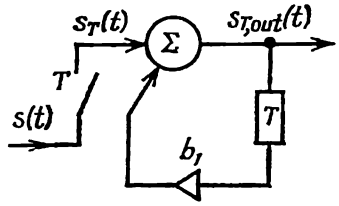


Fig. 13.15. First-order recoursive filter

No matter what the sign of b_1, the condition $|b_1| < 1$ *must be satisfied* to provide for the stability of this circuit. The stability criteria of analog linear circuits with feedback loops, when slightly

modified, are applicable to discrete systems as well.

The amplitude-frequency characteristic and the phase-frequency characteristic of the circuit in question are respectively

$$K_T(\omega)=\frac{1}{\sqrt{1+b_1^2-2b_1\cos\omega T}}=\frac{1}{\sqrt{1+b_1^2-2b_1\cos\left(2\pi\frac{\omega}{\omega_1}\right)}} \quad (13.43)$$

$$\varphi(\omega)=-\arctan\frac{b_1\sin\omega T}{1-b_1\cos\omega T}=\arctan\frac{-b_1\sin\left(2\pi\frac{\omega}{\omega_1}\right)}{1-b_1\cos\left(2\pi\frac{\omega}{\omega_1}\right)} \quad (13.44)$$

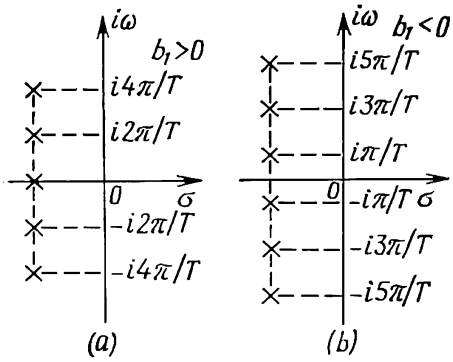


Fig. 13.16. Location of the poles of the transfer function of a recoursive filter

The amplitude-frequency characteristic for several values of b_1 is shown in Fig. 13.17.

The circuit in Fig. 13.15 corresponds to a comb filter separating oscillations of frequencies $\omega = 0,\ \omega_1,\ 2\omega_1,\ \ldots$ when the coefficient b_1 is close to unity, and those of frequencies $\omega = 0.5\omega_1, 1.5\omega_1, 2.5\omega_1, \ldots$ when b_1 is close to -1 (in both cases, $|b_1| < 1$).

The use of a transfer function in the form of (13.40) for the analysis of discrete circuits of higher order is difficult. The analysis can be considerably simplified by using the z-transform method.

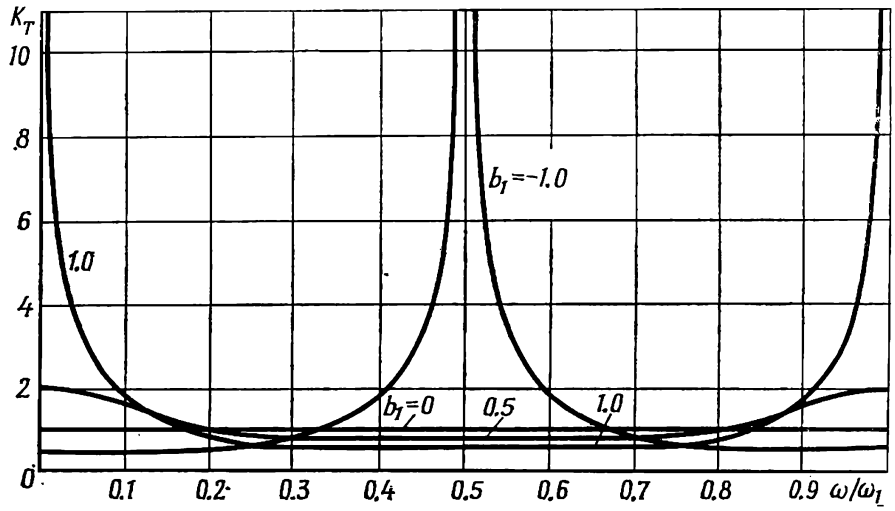


Fig. 13.17. Amplitude-frequency characteristics of a recoursive filter (see Fig. 13.15)

13.8. USE OF THE z-TRANSFORM METHOD FOR THE ANALYSIS OF DISCRETE SIGNALS AND CIRCUITS

The function e^{pT} plays an important role in mathematical description of discrete trains of pulses and discrete circuits. The Laplace transforms of time-dependent processes, as well as circuit transfer

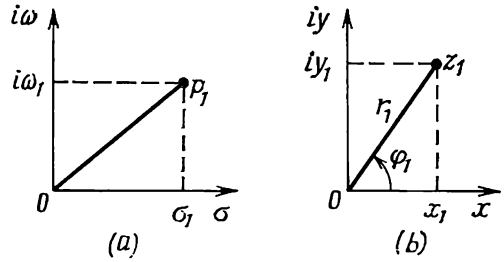


Fig. 13.18. Relation between the coordinates of the point on (*a*) the *p*-plane and (*b*) the *z*-plane

functions which include the function e^{pT}, turn out to be transcedental functions of p, this materially complicating the analysis. It can be simplified by using a new variable z related to p by the relation

$$z = e^{pT}, \quad p = \frac{1}{T} \ln z \qquad (13.45)$$

With such a replacement, the above-mentioned functions of p are transformed into rational functions of the variable z, so that their representation on the plane z is simplified.

The conversion of the plane $p = \sigma + i\omega$ into the plane $z = x + iy$ can be effected by means of the following relations that relate the coordinates σ_1, ω_1 of any point p_1 on the plane p to the coordinates x_1, y_1 of the corresponding point z_1 on the plane z (Fig. 13.18):

$$z_1 = x_1 + iy_1 = e^{(\sigma_1 + i\omega_1)T}$$

$$x_1 = e^{\sigma_1 T} \cos \omega_1 T;$$

$$y_1 = e^{\sigma_1 T} \sin \omega_1 T \qquad (13.46)$$

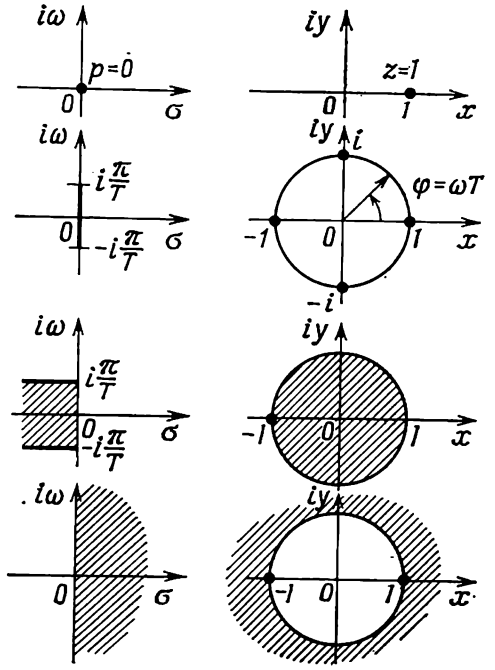


Fig. 13.19. Mapping of points and zones of the *p*-plane into the *z*-plane

In the polar coordinates on the plane z

$$r_1 = |z_1| = \sqrt{x_1^2 + y_1^2} = e^{\sigma_1 T}, \quad \varphi_1 = \arg z_1 = \omega_1 T + m2\pi \qquad (13.47)$$

where m is any integer.

Figure 13.19 shows the mapping of some typical points and regions from the plane p onto the plane z. The point $p = 0$ transforms into the point $z = 1$ on the real axis of the z-plane. As a point of the p-plane moves along the axis $i\omega$ (i.e., at $\sigma = 0$), the corresponding point of the z-plane describes a circle of a unit radius. One complete revolution of the radius-vector corresponds to a change of the frequency ω within the interval $\omega_1 \leqslant \omega \leqslant \omega_1 + 2\pi/T$.

As the point p_1 moves along the axis $i\omega$ within the range from $-i\infty$ to $i\infty$, the point z_1 describes an infinitely large number of circles. Thus, the univalent mapping of p on z exists only for the p-plane band within $\pm i\pi/T$. Inside this band, the left-hand half-plane is projected into the unit circle. All parallel bands of the same width correspond to the same circle. The right-hand half-plane p is converted into the whole z-plane excluding the unit circle.

13.9. z TRANSFORM OF TIME FUNCTIONS

Using the discrete Laplace transforms given in Sec. 13.5, let us set up similar expressions for the z transform by substituting $e^{pT} = z$.

Expression (13.15) assumes the form

$$\hat{S}(z) = S_T(p)_{p=\frac{1}{T}\ln z} = \sum_{k=0}^{\infty} s(kT)\, z^{-k} \tag{13.48}$$

which is called the *direct* z transform.

Let us find the function $\hat{S}(z)$ for some simple time functions $s_T(t)$.

1. A sequence of samples of the signal $s(t) = 1$, $t \geqslant 0$. In this case, $s(kT) = 1$, $k = 0, 1, 2, \ldots, \infty$ and, in accordance with (13.48),

$$\hat{S}(z) = \sum_{k=0}^{\infty} s(kT)\, z^{-k} = \sum_{k=0}^{\infty} z^{-k} = \frac{1}{1-z^{-1}} = \frac{z}{z-1} \tag{13.49}$$

The zero $z_0 = 0$, the pole $z_p = 1$.

2. A sequence of samples of the signal $s(t) = e^{-\alpha t}$, $t \geqslant 0$. In this case, $s(kT) = e^{-\alpha kT}$ and

$$\hat{S}(z) = \sum_{k=0}^{\infty} e^{-\alpha kT} z^{-k} = \sum_{k=0}^{\infty} (e^{-\alpha T} z^{-1})^k = \frac{1}{1-e^{-\alpha T} z^{-1}} = \frac{z}{z-e^{-\alpha T}} \tag{13.50}$$

The zero $z_0 = 0$, the pole $z_p = e^{-\alpha T}$.

3. A sequence of samples of the signal $s(t) = a^{\alpha t}$, $t \geqslant 0$, $a < 1$. In this case, $s(kT) = a^{\alpha kT}$ and

$$\hat{S}(z) = \sum_{k=0}^{\infty} a^{\alpha kT} z^{-k} = \sum_{k=0}^{\infty} (a^{\alpha T} z^{-1})^k = \frac{1}{1 - a^{\alpha T} z^{-1}} = \frac{z}{z - a^{\alpha T}} \qquad (13.51)$$

The zero $z_0 = 0$, the pole $z_p = a^{\alpha t}$.

4. A sequence of samples of the signal $s(t) = \cos \omega_0 T$, $t \geqslant 0$. In this case, $s(kT) = \frac{1}{2} e^{i\omega_0 kT} + \frac{1}{2} e^{-i\omega_0 kT}$ and

$$\hat{S}(z) = \sum_{k=0}^{\infty} \cos \omega_0 kT z^{-k} = \frac{1}{2} \sum_{k=0}^{\infty} (e^{i\omega_0 T} z^{-1})^k$$

$$+ \frac{1}{2} \sum_{k=0}^{\infty} (e^{-i\omega_0 T} z^{-1})^k = \frac{1}{2} \frac{1}{1 - e^{i\omega_0 T} z^{-1}} + \frac{1}{2} \frac{1}{1 - e^{-i\omega_0 T} z^{-1}}$$

$$= \frac{z^2}{2} \frac{(1 - e^{-i\omega_0 T} z^{-1}) + (1 - e^{i\omega_0 T} z^{-1})}{z^2 - 2z \cos \omega_0 T + 1} = \frac{z(z - \cos \omega_0 T)}{z^2 - 2z \cos \omega_0 T + 1} \qquad (13.52)$$

The zeros: $z_{01} = 0$, $z_{02} = \cos \omega_0 T$; the poles: $z_{p1,\,2} = \cos \omega_0 T \pm$ $\pm i \sin \omega_0 T$.

5. A sequence of samples of the signal $s(t) = \sin \omega_0 t$, $t \geqslant 0$. In this case, $s(kT) = (1/2i) e^{i\omega_0 kT} - (1/2i) e^{-i\omega_0 kT}$ and

$$\hat{S}(z) = \sum_{k=0}^{\infty} \sin \omega_0 kT z^{-k} = \frac{1}{2i} \sum_{k=0}^{\infty} (e^{i\omega_0 T} z^{-1})^k - \frac{1}{2i} \sum_{k=0}^{\infty} (e^{i\omega_0 T} z^{-1})^k$$

$$= \frac{z \sin \omega_0 T}{z^2 - 2z \cos \omega_0 T + 1} \qquad (13.53)$$

The zero $z_0 = 0$; the poles: $z_{p1,\,2} = \cos \omega_0 T \pm i \sin \omega_0 T$.

The position of the zeros and poles for the above five signals is shown in Fig. 13.20.

The determination of the original, i.e., the function $s_T(t)$, by the given transform $\hat{S}_T(z)$ is effected by means of the inverse z transform which is obtained by substituting $e^{pT} = z$ into expression (13.16).

Taking into account the relation $Te^{pT} dp = dz$, this expression is reduced to the form

$$s_T(t) = \frac{1}{T} \frac{1}{2\pi i} \oint_{|z| = e^{cT}} \hat{S}(z) z^{(t/T - 1)} dz \qquad (13.54)$$

The integration is round the circle of radius $r = e^{cT}$, which is the transform of the straight line $\sigma = c$ from the plane $p = \sigma + i\omega$. The constant c is determined from the condition that all the poles of the integrand are within the circle of radius $r = e^{cT}$. The circuit

is in the positive direction (i.e., counterclockwise). As frequency changes from $-\infty$ to ∞, the integration contour is circuited infinitely.

By analogy with expression (13.19), the value of the pulse $s(nT)$ at the point $t = nT$ [without the factor $\delta(t - nT)$] can be determined by a simpler expression

$$s(nT) = T\frac{1}{2\pi i}\oint\limits_{|z|=e^{cT}} \hat{S}(z)\, z^n \frac{dz}{Tz} = \frac{1}{2\pi i}\oint\limits_{|z|=e^{cT}} \hat{S}(z)\, z^{(n-1)}\, dz \qquad (13.55)$$

in which a *single circuit* round the integration contour is implied.

In the above examples of functions $\hat{S}(z)$ having their poles on the circle of unit radius [for $s(kT) = 1$, $\cos \omega_0 kT$, $\sin \omega_0 kT$], the con-

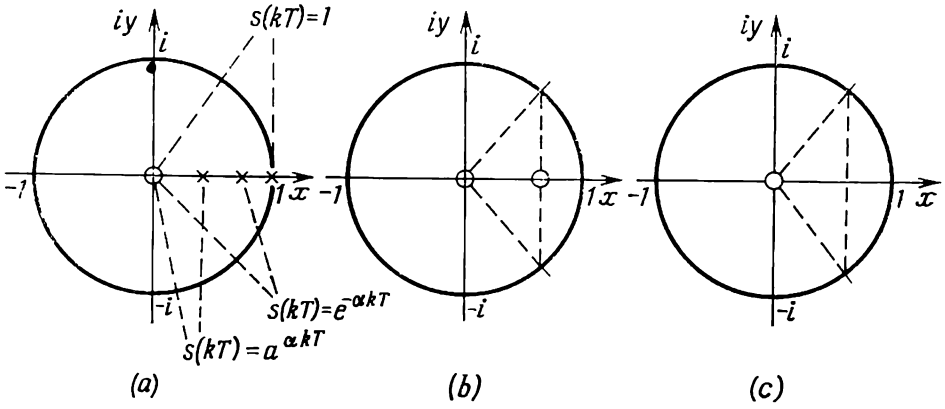


Fig. 13.20. Position of zeros and poles on the z-plane for (*a*) $s(kT) = 1$, $e^{-\alpha kT}$, and $ae^{\alpha kT}$; (*b*) $s(kT) = \cos \omega_0 kT$; and (*c*) $s(kT) = \sin \omega_0 kT$

stant $c > 0$ can be infinitely small. Therefore, the integration contour can be reduced to the circle of radius $r = 1$, with the poles being circuited *outside the circle*, as in the case of integration along the axis $i\omega$ on the plane $p = \sigma + i\omega$, the poles lying on this axis being followed to its *right*.

Taking this condition into account, expressions (13.54) and (13.55) in the text below will be written in one of the following two forms:

$$s_T(t) = \frac{1}{T}\frac{1}{2\pi i}\oint\limits_{|z|=1} \hat{S}(z)\, z^{(t/T-1)}\, dz \qquad (13.56)$$

— with infinite number of circuits round the circle of unit radius;

$$s(nT) = \frac{1}{2\pi i}\oint\limits_{|z|=1} \hat{S}(z)\, z^{(n-1)}\, dz \qquad (13.57)$$

— with a single circuit round the circle of unit radius.

The integration round the circle $r > 1$ will further be excluded from the analysis, since the position of the poles of the function $\hat{S}(z)$ outside the circle $r > 1$ corresponds to infinitely increasing time sequencies that have no physical meaning.

When using expression (13.56), one should take into account a useful relation

$$\frac{1}{T}\frac{1}{2\pi i}\oint\limits_{|z|=1} z^{(t/T-1)}\,dz = \frac{1}{T}\frac{1}{2\pi i}\int\limits_{-i\infty}^{i\infty} e^{pt\,(t/T-1)T}\,e^{pT}\,dp$$

$$= \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{i\omega t}\,d\omega = \delta(t) \qquad (13.58)$$

[see expression (3.17)].

In conclusion, let us consider a *bilateral* z transform which is obtained by substituting e^{pT} into (13.23):

$$\hat{S}(z) = \sum_{k=0}^{\infty} s(kT)\,z^{-k} + \sum_{m=0}^{\infty} s(-mT)\,z^{m} - s(0)$$

$$= \hat{S}_{+}(z) + \hat{S}_{-}(z) - s(0) \qquad (13.59)$$

If the function $s(t)$ is even, the second term on the right-hand side of (13.59) can be reduced to

$$\hat{S}_{-}(z) = \sum_{m=0}^{\infty} s(-mT)\,z^{m} = \sum_{m=0}^{\infty} s(mT)\,z^{m}$$

$$= \sum_{m=0}^{\infty} s(mT)\left(\frac{1}{z}\right)^{-m} = \hat{S}_{+}\left(\frac{1}{z}\right)$$

Thus, with even function $s(t)$, expression (13.59) transforms into

$$\hat{S}(z) = \hat{S}_{+}(z) + \hat{S}_{+}\left(\frac{1}{z}\right) - s(0) = \sum_{k=-\infty}^{\infty} s(kT)z^{-k} \qquad (13.60)$$

In this expression, $\hat{S}_{+}$ stands for unilateral transform.

The inverse z transformation is performed by means of expressions (13.56) and (13.57).

13.10. z TRANSFORMS OF TRANSFER FUNCTIONS OF DISCRETE CIRCUITS

Let us apply the z transformation to the transfer function of a discrete circuit. Substituting $e^{pT} = z$ into expression (13.40) gives

$$\hat{K}(z) = \hat{S}_{out}(z)/\hat{S}(z) = \sum_{k=0}^{H} a_k z^{-k} \Big/ \left(1 - \sum_{k=1}^{M} b_k z^{-k}\right) \qquad (13.61)$$

From this expression it is clear that the transfer function of a discrete filter is fractionally rational. If expression (13.61) is specified in advance, one can easily set up an equation in the from of (13.38), defining the algorithm of conversion of the input pulse train into the output one. For this purpose, it is sufficient to provide the terms of the form $s[(n-k)T]$ in equation (13.38) with the coefficients a_k multiplying the terms z^{-k} in the numerator of expression (13.61) and to provide the terms of the form $s_{out}[(n-k)T]$ in equation (13.38) with the coefficients b_k multiplying the terms z^{-k} in the denominator of expression (13.61).

However, it should be noted that not every fractionally rational function can be realized in the form of a transfer function of a filter. For example, let a transfer function be specified in the form of a ratio of polynomials in positive powers of z:

$$\hat{\mathbf{K}}(z)=\frac{a_0z^H+a_1z^{H-1}+\ldots+a_H}{b_0z^M-b_1z^{M-1}-\ldots-b_M} \tag{13.62}$$

Dividing the numerator and denominator by b_0z^M, we reduce this expression to the form

$$\hat{\mathbf{K}}(z)=\frac{\frac{a_0}{b_0}z^{H-M}+\frac{a_1}{b_0}z^{H-M-1}+\ldots+\frac{a_H}{b_0}z^{-M}}{1-\frac{b_1}{b_0}z^{-1}-\ldots-\frac{b_M}{b_0}z^{-M}}$$

If $H>M$, the first term in the numerator (with a positive exponent of z) forms in equation (13.38) a term of the form $(a_0/b_0)\times$ $\times s[(n+k)T]$, where $k=H-M>0$ corresponds to the pulse $s(n+k)$ leading the input pulse $s(n)$, which, of course, is impossible. From this it follows that the filter can be realized, provided that the *power of the denominator in (13.62) is equal to or higher than the power of the numerator.*

Taking this into account, let us write the transfer function in the following equivalent forms:

$$\hat{\mathbf{K}}(z)=\frac{a_0+a_1z^{-1}+a_2z^{-2}+\ldots+a_Hz^{-H}}{1-b_1z^{-1}+b_2z^{-2}-\ldots-b_Mz^{-M}}$$

$$=\frac{z^{M-H}(a_0z^H+a_1z^{H-1}+a_2z^{H-2}+\ldots+a_H)}{z^M-b_1z^{M-1}-b_2z^{M-2}-\ldots-b_M} \tag{13.63}$$

$$\hat{\mathbf{K}}(z)=a_0z^{M-H}\frac{(z-z_{01})(z-z_{02})\ldots(z-z_{0H})}{(z-z_{p1})(z-z_{p2})\ldots(z-z_{pM})} \tag{13.64}$$

In expression (13.63), the coefficients a_k and b_k should be taken with the same signs as they enter into (13.38).

In expression (13.64), z_{0n} are the zeros, and z_{pn}, the poles of the transfer function; z_{0n} and z_{pn} may be either real or complex numbers. In the first case they are located on the real axis, while in the second, they form complex-conjugate pairs.

The zeros can be located at any point of the plane z, while the poles, only *inside the circle of unit radius.* This condition stems from the requirement for the stability of the circuit; when analyzing the behaviour of the transfer function on the plane p, the stability condition requires that the poles should be located in the left-hand half-plane. As mentioned above, the left-hand half-plane p is projected into the unit circle on the plane z.

To transform from the function $\hat{\mathbf{K}}(z)$ to the function $\mathbf{K}_T(i\omega)$, one should, as it follows from (13.46), set $z = e^{i\omega T}$ ($\sigma = 0$)

Thus,

$$\mathbf{K}_T(i\omega) = \frac{a_0 + a_1 e^{-i\omega T} + a_2 e^{-i2\omega T} + \ldots + a_H e^{-iH\omega T}}{1 - b_1 e^{-i\omega T} - b_2 e^{-i2\omega T} - \ldots - b_M e^{-iM\omega T}}$$

$$= a_0 \frac{(e^{i\omega T} - z_{01})(e^{i\omega T} - z_{02}) \ldots (e^{i\omega T} - z_{0H})}{(e^{i\omega T} - z_{p1})(e^{i\omega T} - z_{p2}) \ldots (e^{i\omega T} - z_{pM})} \qquad (13.65)$$

To determine the amplitude-frequency characteristic of the circuit within the range $(0, 2\pi/T)$, it is necessary to calculate the modulus of expression (13.65) with ωT changing from 0 to 2π, i.e., with the circle of unit radius on the z-plane being circuited only once. As the circle is being circuited further, the amplitude-frequency characteristic is periodically repeated.

The magnitudes of the differences $e^{i\omega T} - z_{0k}$ and $e^{i\omega T} - z_{pk}$ are the distances from the point of the circle corresponding to the angle ωT to the zero z_{0k} or pole z_{pk}. Denoting these distances as R_{0k} and R_{pk}, we obtain the following formula for the amplitude-frequency characteristic:

$$K_T(\omega) = a_0 \frac{R_{01} R_{02} \ldots R_{0H}}{R_{p1} R_{p2} \ldots R_{pM}} \qquad (13.66)$$

convenient for graphical calculations.

The calculations are particularly simplified when the amplitude-frequency characteristic is plotted to a logarithmic scale:

$$K_T(\omega)_{\text{dB}} = 20 \left[\log_{10} a_0 + \sum_{k=1}^{H} \log_{10} R_{0k} - \sum_{k=1}^{M} \log_{10} R_{pk}\right] \qquad (13.67)$$

If the zeros and poles of the transfer function are known, one can readily calculate the coefficients a_k and b_k using well-known algebraic relations. The calculation of the zeros and poles by the given coefficients a_k and b_k is a more complicated problem (for $M > 2$).

The transfer function $\hat{\mathbf{K}}(z)$ and the impulse response $g_T(t)$ are interrelated by a pair of z transforms stemming directly from expressions (13.48), (13.56), and (13.57), when in the first of them $s(kT)$ is replaced by $g(kT)$ and in (13.56) and (13.57) $\hat{\mathbf{S}}(z)$ is replaced by $\hat{\mathbf{K}}(z)$:

$$\hat{\mathbf{K}}(z) = \sum_{k=0}^{\infty} g(kT)\, z^{-k} \tag{13.68}$$

$$g_T(t) = \frac{1}{T}\frac{1}{2\pi i}\oint \hat{\mathbf{K}}(z)\, z^{(t/T-1)}\, dz \tag{13.69}$$

(infinite number of circuits round the integration contour)

$$g(nT) = \frac{1}{2\pi i}\oint \hat{\mathbf{K}}(z)\, z^{(n-1)}\, dz \tag{13.70}$$

(a single circuit round the integration contour).

13.11. EXAMPLES OF THE ANALYSIS OF DISCRETE FILTERS, BASED ON THE z-TRANSFORM METHOD

13.11-1. Second-Order Nonrecoursive Filter

In accordance with expression (13.30), for $H = 2$ the transfer function of the filter shown in Fig. 13.21*a* is transformed into the

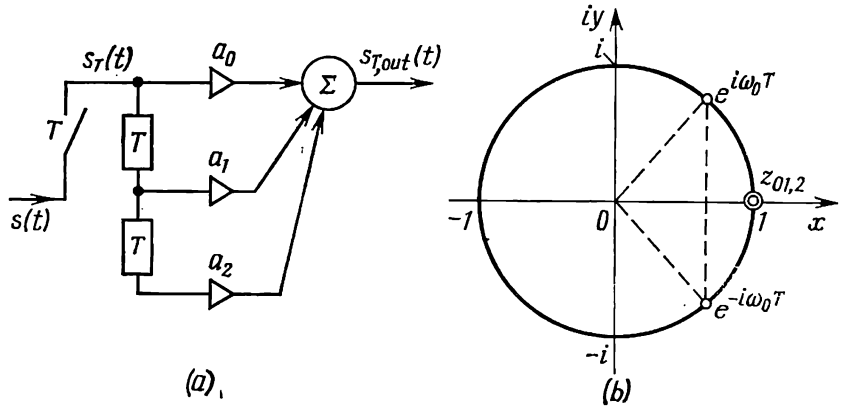


Fig. 13.21. (*a*) Second-order digital filter and (*b*) the location of zeros on the z-plane

function

$$\hat{\mathbf{K}}(z) = \sum_{k=0}^{H} a_k z^{-k} = a_0 + a_1 z^{-1} + a_2 z^{-2} = a_0 \frac{z^2 + \frac{a_1}{a_0} z + \frac{a_2}{a_0}}{z^2} \tag{13.71}$$

This function has zeros at the points

$$z_{01,2} = -\frac{1}{2}\frac{a_1}{a_0} \pm \sqrt{\frac{1}{4}\left(\frac{a_1}{a_0}\right)^2 - \frac{a^2}{a_0}}$$

The double pole at the point $z_n = 0$ has no effect on the behaviour of the transfer function.

Of special interest are the cases $a_0 = a_2 = 1$ and $|a_1| < 2$, where

$$z_{01,2} = -a_1/2 \pm i\sqrt{1 - a_1^2/4}$$

The modulus of this expression is equal to unity, so that the complex-conjugate zeros lie on the unit-radius circle. In particular, for $a_1 = -2$, the double zero is located at the point $z = 1$ (Fig. 13.21*b*). This case corresponds to the second-order rejector filter with infinitely high attenuation at the frequency $\omega = 0$ ($e^{i\omega T} = 1$), which is widely used in practice.

The amplitude-frequency characteristic of such a filter can easily be found from expression (13.71) by substituting $a_0 = a_2 = 1$, $z^{-1} = e^{-i\omega T}$ and multiplying the right-hand side by $|e^{i\omega T}| = 1$:

$$|\hat{K}(z)| = |\hat{K}(e^{i\omega T})| = |1 - 2e^{-i\omega T} + e^{-i2\omega T}| = |e^{i\omega T} - 2 + e^{-i\omega T}| = 2|\cos\omega T - 1| = 4\sin^2\frac{\omega T}{2} \qquad (13.72)$$

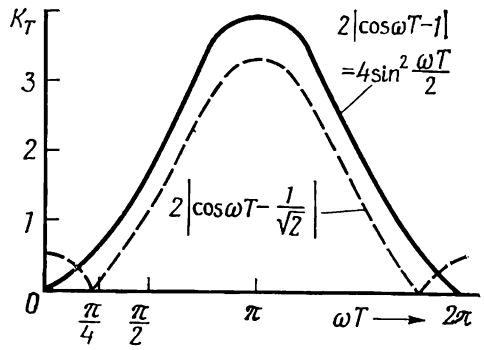


Fig. 13.22. Amplitude-frequency characteristic of a filter (Fig. 13.21*a*) for $a_0 = 1$ and $a_1 = -1$

The graph of this function is shown in Fig. 13.22 (continuous line). Comparison between (13.72) and (13.35) shows that this second-order filter is equivalent to the cascade connection of two first-order filters with coefficients $a_0 = 1$ and $a_1 = -1$.

By changing the coefficient a_1, it is possible to shift the zeros $z_{01,2}$ along the unit-radius circle, and this is equivalent to shifting the zero of the amplitude-frequency characteristic along the frequency axis. In particular, for $a_1 = -\sqrt{2}$, $z_{01,2} = (1/\sqrt{2})(1 \pm i) = e^{\pm i\omega_0 T}$, $\omega_0 T = 45°$ (Fig. 13.21*b*). The respective amplitude-frequency characteristic is shown in Fig. 13.22 by a dashed line

13.11-2. First-Order Recoursive Filter (Fig. 13.15)

The transfer function (13.42) is transformed into

$$\hat{K}(z) = 1/(1 - b_1 z^{-1}) = z/(z - b_1) \qquad (13.73)$$

This function has its zero at the point $z_0 = 0$ and the pole at the point $z_p = b_1$.

Let us find the impulse response of the filter by means of formula (13.69). Representing $\hat{K}(z)$ in the form of a geometric progression

$$\hat{K}(z) = 1 + b_1 z^{-1} + b_1^2 z^{-2} + \ldots$$

and applying expression (13.69) to each term, we obtain

$$g_T(t) = \frac{1}{T}\frac{1}{2\pi i}\oint\limits_{|z|=1} z^{(t/T-1)}\,dz + \frac{b_1}{T}\frac{1}{2\pi i}\oint\limits_{|z|=1} z^{(t/T-2)}\,dz + \ldots$$

$$= \delta(t) + b_1\delta(t-T) + b_1^2\delta(t-2T) + \ldots$$

$$= \sum_{k=0}^{\infty} b_1^k \delta(t-kT), \quad g(kT) = b_1^k \qquad (13.74)$$

From this example one can clearly see the advantage of the recoursive filter over the nonrecoursive one. To obtain the above impulse response, only one memory element T is required in the case of the recoursive filter, while for the nonrecoursive one, it is necessary to have quite a number of these (theoretically an infinite number). In the recoursive filter, this advantage is realized owing to the fact that the pulse is made to circulate through the feedback loop with a delay T.

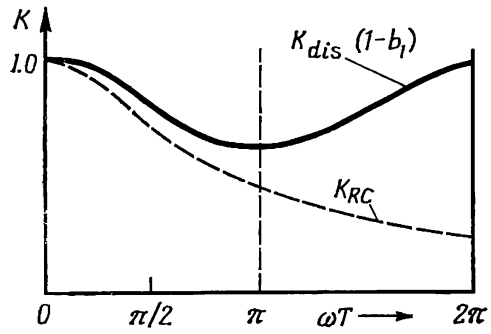


Fig. 13.23. Amplitude-frequency characteristic of a digital filter (continuous line) and an analog R-C circuit (dashed line) having equivalent impulse response characteristics

Representing b_1 in the form $b_1 = e^{-\alpha T}$, let us write expression (13.74) in the form

$$g_T(t) = \sum_{k=0}^{\infty} e^{-\alpha kT}\delta(t-kT)$$

$$= e^{-\alpha t}\sum_{k=0}^{\infty}\delta(t-kT) \qquad (13.75)$$

from which it follows that the discrete impulse response of the circuit under consideration coincides with a sequence of samples of the impulse response of an analog R-C circuit whose time constant meets the requirement

$$b_1 = e^{-\alpha T} = e^{-T/RC}, \text{ or } RC = T/\ln(1/b_1)$$

However, the amplitude-frequency characteristics of the two circuits are essentially different. For the discrete circuit, the amplitude-frequency characteristic is defined by formula (13.43), while for the analog one, it is determined by the expression

$$K_{RC}(\omega) = 1/\sqrt{1+(\omega RC)^2}$$

In Fig. 13.23 comparison is made between the amplitude-frequency characteristic K_{dis} of the discrete circuit (normalized by the maximum value) for $b_1 = 0.2$ and the amplitude-frequency characteristic K_{RC} [for $RC = T/\ln(1/b_1)$].

The deformation of the amplitude-frequency characteristic of the discrete circuit is due to the comb structure of the transfer function (see Sec. 13.5) and the superposition of the tails of the characteristic belonging to adjacent frequency intervals.

13.11-3. Second-Order Recoursive Filter (Fig. 13.24)

First, let us write the transfer function in the form

$$\hat{\mathbf{K}}(z) = 1/(1 - b_1 z^{-1} - b_2 z^{-2}) = z^2/(z^2 - b_1 z - b_2) = z^2/(z - z_{p1})(z - z_{p2}) \tag{13.76}$$

corresponding to the case $a_0 = 1$, $a_1 = 0$, $a_2 = 0$ where the zeros of the transfer function (in the given case, the double zero) are only at

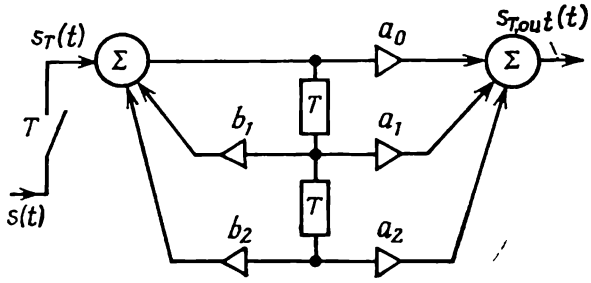


Fig. 13.24. Second-order recoursive digital filter

the point $z = 0$, i.e., at the centre of the unit-radius circle.

The roots of the equation $z^2 - b_1 z - b_2 = 0$ (the poles) are

$$z_{p1,2} = b_1/2 \pm \sqrt{b_1^2/4 + b_2} \tag{13.77}$$

For $b_2 < 0$ and, in addition, $|b_2| > b_1^2/4$, the poles z_{p1} and z_{p2} are complex-conjugate numbers:

$$z_{p1} = b_1/2 + i\sqrt{|b_2| - b_1^2/4}, \quad z_{p2} = z_{p1}^*$$

In this case

$$(z - z_{p1})(z - z_{p2}) = z^2 - 2\mathrm{Re}(z_{p1,2})\, z + |z_{p1,2}|^2$$

from which stem the following relations between the coefficients in the polynomial in (13.76) and the poles $z_{p1,2}$:

$$b_1 = 2\mathrm{Re}(z_{p1,2}), \quad b_2 = -|z_{p1,2}|^2$$

Representing $z_{p1,2}$ in the form

$$z_{p1,2} = |z_{p1,2}|\, e^{\pm i\omega_p T} = re^{\pm i\varphi_p} \tag{13.78}$$

where $r = |z_{p1,2}|$ is the distance to the pole from the origin of coordinates and $\varphi_p = \omega_p T$ is the pole azimuth (Fig. 13.25), we obtain

$$b_1 = 2r \cos \omega_p T, \quad b_2 = -r^2 \tag{13.79}$$

To determine the amplitude-frequency characteristic of the circuit in question, let us substitute $z = e^{i\omega T}$ into (13.76) and take the modulus

$$|\hat{K}(e^{i\omega T})| = K_T(\omega) = 1/|(e^{i\omega T} - re^{i\omega_p T})(e^{i\omega T} - re^{-i\omega_p T})| \tag{13.80}$$

With the given position of the poles (i.e., with r and $\omega_p t$ being specified), the amplitude-frequency characteristic is preferably plot-

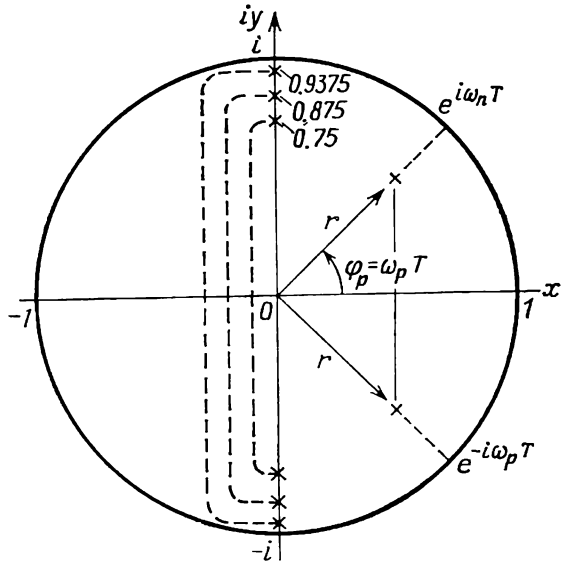


Fig. 13.25. Position of the poles of a second-order digital filter on the z-plane

ted by using formula (13.66), measuring R_{p1} and R_{p2} in the drawing. In the given case, in order to simplify the calculations, we shall use formula (13.80) for the specific case $\omega_p T = 90°$. In this case expression (13.80) is easily reduced to the form

$$K_T(\omega T) = 1/\sqrt{1 + 2r^2 \cos 2\omega t + r^4} \tag{13.81}$$

The graphs of the function $K_T(\omega T)$ for $r = 0.75$, 0.875, and 0.9375 are shown in Fig. 13.26. As r approaches unity, the circuit in question approaches a resonator with a very high Q-factor. However, in this case the circuit stability may be disturbed.

Now let us consider a second-order transfer function of a more general form corresponding to the circuit in Fig. 13.24:

$$\hat{K}(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{1 - b_1 z^{-1} - b_2 z^{-2}} = a_0 \frac{(z - z_{01})(z - z_{02})}{(z - z_{p1})(z - z_{p2})} = \frac{\alpha_T(z)}{\beta_T(z)} \tag{13.82}$$

As mentioned in Sec. 13.6 [see formula (13.40) and explanations to this formula], the filter with transfer function (13.82) may be treated as a cascade connection of a nonrecoursive filter [with a transfer

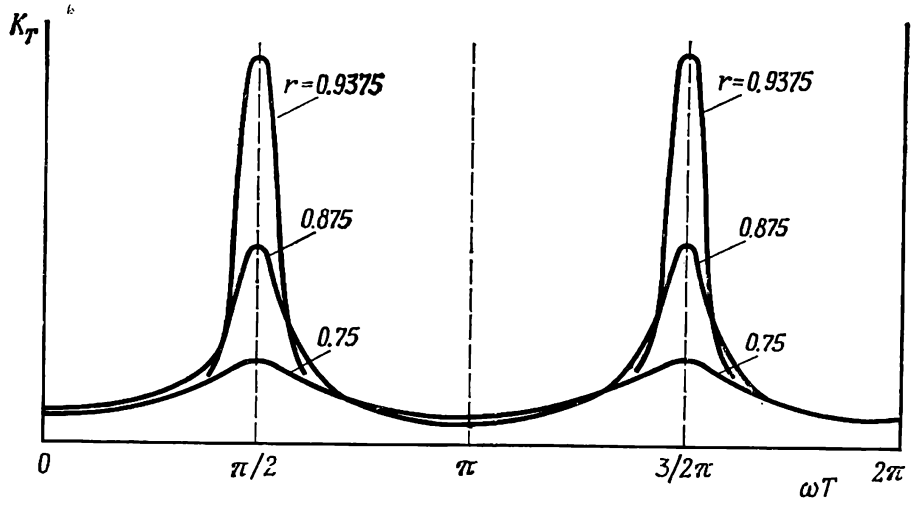


Fig. 13.26. Amplitude-frequency characteristics of a second-order recoursive filter (see Figs. 13.24 and 13.25)

function $\alpha_T(z)$] and a recoursive one [with a transfer function $1/\beta_T(z)$]. Such a combination can be used, for example, in a rejec-

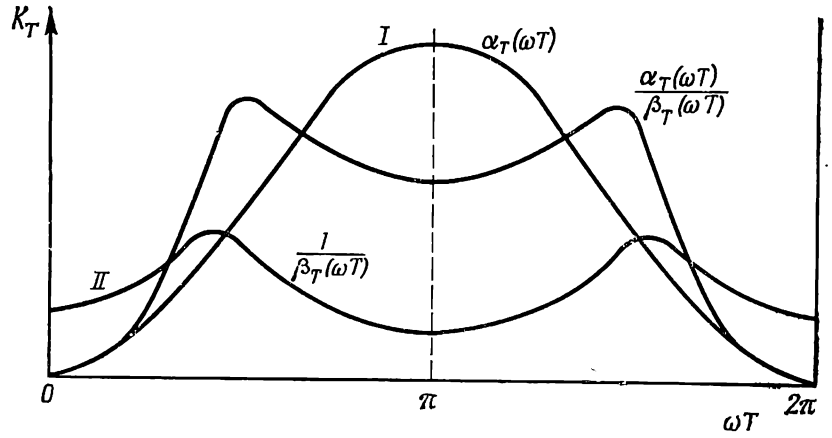


Fig. 13.27. Amplitude-frequency characteristics of (*I*) a recoursive link with feedforward loops, (*II*) a link with feedback loops, and a digital filter as a whole

tion filter like the one discussed in Sec. 13.11-1, but provided with additional feedback circuits for levelling out the amplitude-frequency characteristic within the passband of the filter.

Shown in Fig. 13.27 are the graph of the function $|\alpha_T(\omega T)|$, transferred from Fig. 13.22 (for $a_0 = a_2 = 1$, $a_1 = -2$), that of the function $|1/\beta_T(\omega T)|$ for $b_1 = 0.21875$ and $b_2 = 0.4375$, and the resultant amplitude-frequency characteristic.

13.12. ANALOG-TO-DIGITAL CONVERSION. QUANTIZATION NOISE

In the preceding chapters, when studying discrete filters, the question of unavoidable error of conversion of the input signal from analog into digital form was not discussed. This error arises when the signal is quantized into a finite, limited number of levels. To reveal the nature of this error, let us return to the block diagram of a system for digital processing of a signal, shown in Fig. 13.1, and isolate two devices from this system: the analog-to-digital converter (*ADC*) and the digital-to-analog converter (*DAC*).

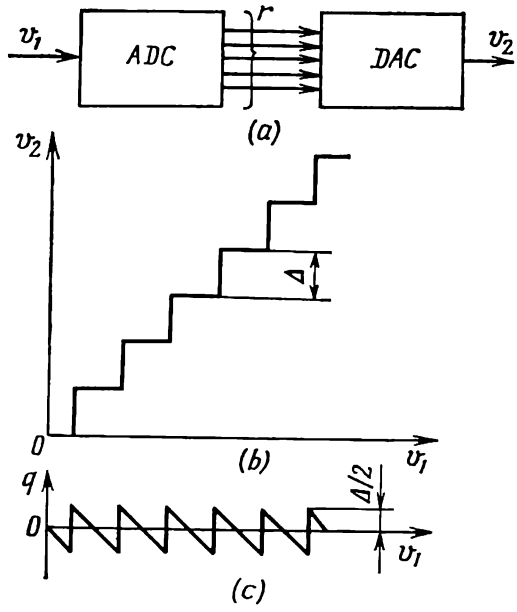


Fig. 13.28. (*a*) Analog-to-digital and digital-to-analog conversion, (*b*) quantization characteristic, and (*c*) quantization error

Let us first consider the joint operation of these devices, without taking into account the digital filter (Fig. 13.1), when a d-c voltage v_1 of different levels is applied to the input of the *ADC* (Fig. 13.28*a*). The main parameter of the *ADC* is the number of digits used for coding the input voltage. With a binary code, the number of digits is defined by the number of the binary elements (e.g., flip-flops) used, each of which can be in one of the two states: with zero or nonzero voltage at the output. The two states that a flip-flop can assume are customarily labelled as a binary 0 and a binary 1. With r binary elements, we have a combination (coded word) of r symbols at the *ADC* output, each symbol being capable of assuming one of two values (0 or 1).

As mentioned in Sec. 13.1, the number of possible combinations $L = 2^r$ determines the number of discrete levels, into which the range of change of the input voltage v_1 can be divided.

The inversion is effected in the *DAC*. Each combination of binaries 0 and 1 fed to the *DAC* input corresponds to a definite discrete level of the output voltage v_2. As a result, with a constant quantizing step Δ, the dependence of v_2 on v_1 assumes the form of a broken line shown in Fig. 13.28*b*.

The device shown in Fig. 13.28*a* and having such a characteristic should be regarded as nonlinear, and the difference $v_2 - v_1 = q$,

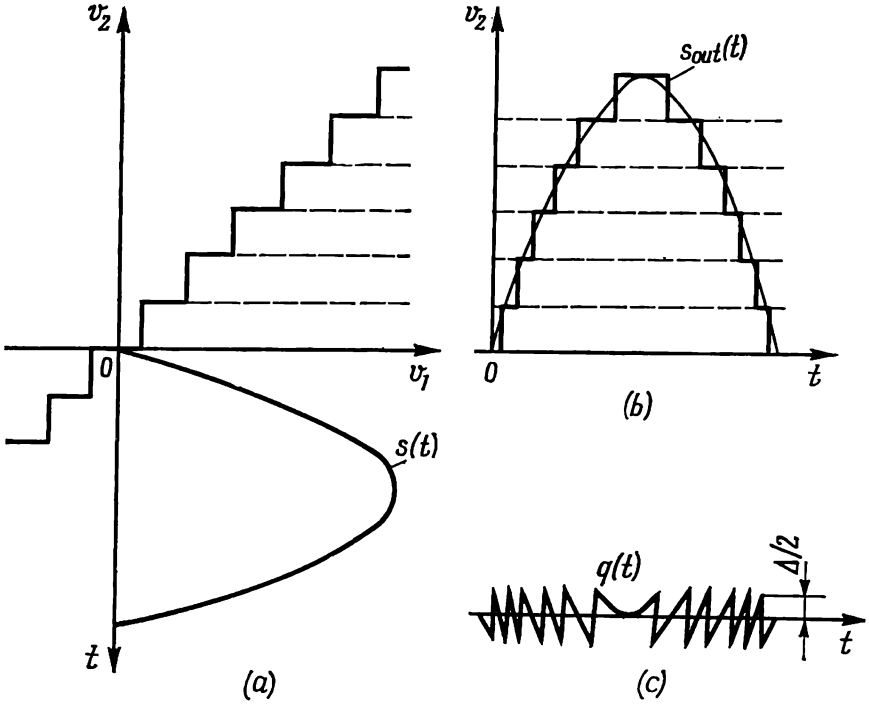


Fig. 13.29. Signal at (*a*) the input and (*b*) the output of a quantizing device, and (*c*) quantization noise

as the quantization error. It is clear that the maximum error, whose magnitude does not exceed $\Delta/2$, remains unchanged with an increase in v_1 (Fig. 13.28*c*).

Let us extend this study to cover a harmonic input oscillation $s(t)$ (Fig. 13.29*a*). The oscillation $s_{out}(t)$ assumes a stepped shape differing from the input oscillation $s(t)$ (shown by a thin line in Fig. 13.29*b*) and the quantization error assumes the form of the function

$$q(t) = s_{out}(t) - s(t) \tag{13.83}$$

shown in Fig. 13.29*c*.

When the amplitude and phase of the harmonic oscillation $s(t)$ change within a wide range, only the repetition frequency of the teeth changes; their shape remains close to triangular at a constant

amplitude $\Delta/2$. The function $q(t)$ can be referred to as the *quantization noise*. The average power of the quantization noise can easily be calculated. Assuming that the teeth have a triangular shape (Fig. 13.29*c*) with an amplitude $\Delta/2$, the average power over the duration of one tooth is equal to $\frac{1}{3}(\Delta/2)^2 = \Delta^2/12$. Since this quantity is independent of the tooth duration, we may consider that the average power of the quantization noise

$$P_q = \Delta^2/12 \tag{13.84}$$

This result, derived for a harmonic signal, can be extended to cover any other signal, including a random one. The only difference is that the function $q(t)$ will be a random process due to the random character of the tooth duration.

It is not difficult to calculate the signal-to-noise ratio in quantization. With the height of the quantizing step equal to Δ and the total number of the steps falling within the characteristic of the *ADC* equal to L, the amplitude of the harmonic signal must not exceed $L\Delta/2$ and the average power of the signal, $\frac{1}{2}(L\Delta/2)^2$ (to avoid any limitation of the signal). Consequently, the signal-to-noise ratio in quantizing a harmonic oscillation

$$P_s/P_q \leqslant 3L^2/2 \tag{13.85}$$

Since the number L of levels is related to the number r of binary digits by the relation $L = 2^r$, expression (13.85) can be represented in the form

$$P_s/P_q = (3/2)2^{2r} \tag{13.86}$$

This relation may be regarded as a particular case of the general expression

$$P_s/P_q = 3 \times 2^{2r}/K_{pf}^2 \tag{13.87}$$

where K_{pf} is the peak factor of the signal, i.e., the ratio of its maximum value to the root-mean-square value.

For a harmonic oscillation, $K_{pf} = \sqrt{2}$, and this leads to expression (13.86). For a random signal obeying a normal distribution law, K_{pf} may be taken at 2.5 to 3 (see Sec. 4.2-4); in this case $P_s/P_q \approx \approx 2^{2r}/3$, and the root-mean-square voltage of the signal must not exceed about $L\Delta/6$. The physical meaning of expression (13.87) is obvious: as the number r of digits increases, the number of the discrete levels falling within the given range of $s(t)$ grows very rapidly and consequently, the step Δ between two adjacent levels is reduced.

Rough estimation of the ratio of the signal to the quantization noise is based on the relation $P_s/P_q \approx 2^{2r}$ or, in decibels,

$$D_{\text{dB}} = (P_s/P_q)_{\text{dB}} = 10 \log_{10} 2^{2r} = 10 \times 2r \log_{10} 2 \approx 6r \tag{13.88}$$

In modern analog-to-digital converters, the number of digits may be as great as ten and more. In this case, the value of D_{dB} characterizing the dynamic range of the analog-to-digital converter comes to approximately 60 dB (6 dB per digit)*.

Another important feature of the quantization noise is its spectral characteristic. In the case of a harmonic oscillation, this noise is a periodic function of time. Its spectrum, like that in any other nonlinear conversion, has a line structure comprising only

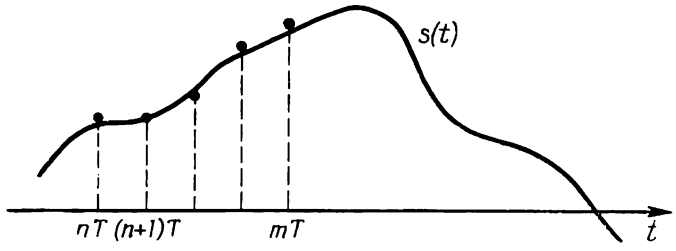


Fig. 13.30. Determining the quantization error

those frequencies that are multiples of the frequency of the input oscillation. The noise spectrum is rich in higher harmonics due to the toothed shape of the function $q(t)$ (Fig. 13.29*c*).

In the case of an input signal in the form of a random process with variance σ_s^2 and root-mean-square width $f_{s,\,rms}$ of the energy spectrum, the statistical characteristics of the quantization noise depend not only on the characteristics of the initial process $s(t)$, but also to a great extent on the relation between σ_s and Δ. In particular for $\sigma_s \gg \Delta$, the width $f_{q,\,rms}$ of the energy spectrum $W_q(\omega)$ of the quantization noise is much wider than $f_{s,\,rms}$ of the process $s(t)$.

Let us introduce into consideration the time quantization (sampling) of the input signal. Shown in Fig. 13.30 are one of the realizations of a random signal $s(t)$ and a set of samples taken at intervals T.

In the *ADC*, each sample is converted into a digital code, as it was described in Sec. 13.1 and at the beginning of this section for a d-c voltage.

As is clear from the preceding arguments the conversion is effected with an error lying within $\pm\,\Delta/2$. If the signal being sampled is random and the change in the function $s(t)$ during time T exceeds one or several Δ, the errors at different sampling points nT and $(n+1)\,T$ may be considered mutually independent and equiprobable. The variance of the random variable q equiprobable within the range $(-\Delta/2,\ \Delta/2)$ is equal to $(1/3)\,(\Delta/2)^2$ (see Sec. 4.2-1). This result coincides with expression (13.84) obtained by averaging the power

* At a peak factor $K_{pf} \approx 3$ [see formula (13.87)] the value of D_{dB} decreases to 5.5 dB per digit.

of quantization noise over time. The above assumptions are equivalent to the statement that the discrete sequence of errors $q(nT)$ corresponds to samples of uncorrelated noise, i.e., noise with a uniform energy spectrum. As mentioned above, this spectrum is much wider than the spectrum of the original random process $s(t)$. In this connection, the quantization noise is usually regarded as white noise, additive with respect to $s(t)$. Since the quantization process is effected at the input of the digital filter, the quantization noise may be treated as natural noise of the digital filter (referred to its input).

Let us find the energy spectrum of the quantization noise. Let the total spectrum width of the quantization noise be equal to $f_{q,\ rms}$ in the absence of sampling. When the quantization noise is sampled at intervals $T = 1/f_1$, the resultant spectrum is a sum of partial spectra shifted with respect to one another by an amount $\omega_1 = 2\pi/T$ (see Sec. 13.3, Fig. 13.6). A specific feature of the spectrum in question is that $f_{q,\ rms} \gg 1/T = f_1$, so that the component spectra overlap many times over.

In the frequency range $(0, f_1)$, each individual spectrum carries a power of $(\Delta^2/12) f_1/f_{q,\ rms}$. But the number of the overlapping spectra is equal to $f_{q,\ rms}/f_1$. Therefore, the total power of the quantization noise within the frequency band $(0, f_1)$ is equal to $\Delta^2/12$. Thus, we may consider that within the above frequency interval the energy spectrum is uniform (white noise) and equal to

$$W_q(\omega) = (\Delta^2/12)\, 1/f_1, \quad 0 < f < f_1 \tag{13.89}$$

With the amplitude-frequency characteristic of the digital filter being $K^2_T(\omega)$, the energy spectrum of the quantization noise at the filter output is

$$W_{q,\ out}(\omega) = (1/12)(\Delta^2/f_1) K^2_T(\omega), \quad -\omega_1/2 < \omega < \omega_1/2 \tag{13.90}$$

and the average power (variance)

$$\sigma^2_{q,\ out} = \frac{1}{12} \frac{\Delta^2}{f_1} \frac{1}{2\pi} \int_{-\omega_1/2}^{\omega_1/2} K^2_T(\omega)\, d\omega \tag{13.91}$$

For quantitative analysis, let us find the main parameters of the quantization noise at the output of the second-order rejection filter discussed in Sec. 13.11-1 and having the following specifications:

— number of quantization digits $r = 8$;
— *ADC* range 10 V;
— sampling interval $T = 1/f_1 = 1$ ms, $f_1 = 1000$ Hz.

The quantization step Δ can be found by dividing 10 V by the number of quantization levels:

$$L = 2^r = 2^8 = 256, \quad \Delta = \frac{10}{256} \approx 0.04 \text{ V} = 40 \text{ mV}$$

The variance of noise at the input

$$\sigma_q^2 \approx \Delta^2/12 = (4 \times 10^{-2})^2/12 \approx 1.3 \times 10^{-4} \text{ V}^2,$$

$$\sigma_q \approx 1.1 \times 10^{-2} \text{ V} = 11 \text{ mV}$$

Using the amplitude-frequency characteristic $K_T(\omega) = 4\sin^2 \frac{\omega T}{2}$ [see formula (13.72)], we find

$$\frac{1}{2\pi} \int_{-\omega_1/2}^{\omega_1/2} K_T^2(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\omega_1/2}^{\omega_1/2} 16 \sin^4 \frac{\omega T}{2}\, d\omega$$

$$= \frac{1}{2\pi} 16 \times 2 \left(\frac{2}{T}\right) \int_0^{\pi/2} \sin^4 x\, dx = \frac{2 \times 3}{T}$$

Now, using formula (13.91), we obtain

$$\sigma_{q,\,out}^2 = \frac{\Delta^2}{12} \frac{1}{f_1} \frac{6}{T} = \frac{\Delta^2}{2}, \quad \sigma_{q,\,out} = \frac{\Delta}{\sqrt{2}} = \frac{40}{\sqrt{2}} \approx 26 \text{ mV}$$

Thus, the level of inherent quantization noise at the output of the filter in question is equal to 26 mV.

The energy spectrum of this noise repeats the shape of the square of the amplitude-frequency characteristic:

$$W_{q,\,out}(\omega) \approx \frac{\Delta^2}{12} \frac{1}{f_1} 16 \sin^4 \left(\frac{\omega T}{2}\right) \approx 2 \times 10^{-6} \sin^4 \left(\frac{\omega T}{2}\right) \text{ V}^2/\text{Hz}$$

In conclusion, let us note the requirements for analog-to-digital converters, depending on the rate of change of the input signal $s(t)$. The sampling duration τ_s is selected so short that the change in $s(t)$ during the time τ_s is negligible. In any case, this change must be less than Δ. In modern analog-to-digital converters τ_s is reduced to nanoseconds.

In Sec. 13.1 it was mentioned that the electronic switch which is used for taking samples of the signal $s(t)$ is provided with an *R-C* circuit for storing the sample level during the time required for the operation of the *ADC*. In high-speed analog-to-digital converters this time comes to tens of nanoseconds.

13.13. DIGITAL-TO-ANALOG CONVERSION AND RESTORATION OF A CONTINUOUS SIGNAL

The inversion of a signal from digital into analog form* is effected by means of two devices: (a) a digital-to-analog converter (*DAC*) and (*b*) a synthesizing filter (see the diagram in Fig. 13.1).

* In many cases of application of digital filters, there is no need in digital-to-analog conversion. For example, in radar systems with digital processing of signals, the latter are fed into the computer directly in digital form.

represented in the form of the convolution of a square pulse $v_0(t)$ shown in Fig. 13.32 with a function $s(kT)\,\delta(t-kT)$.

In fact

$$s(kT)\int_{-\infty}^{\infty} v_0(t-\tau)\,\delta(\tau-kT)\,d\tau$$

$$=s(kT)\,v_0(t-kT)=\begin{cases} s(kT) & \text{for } kT<t<kT+\tau_0 \\ 0 & \text{for } t<kT \text{ and } t>kT+\tau_0 \end{cases}$$

Thus, the entire train of pulses at the DAC output (without digital filter and for the circuit shown in Fig. 13.28*a*) can be written in the form

$$s'_T(t)=\sum_{k=0}^{\infty} s(kT)\int_{-\infty}^{\infty} v_0(t-\tau)\,\delta(\tau-kT)\,d\tau$$

$$=\int_{-\infty}^{\infty} v_0(t-\tau)\Big[\sum_{k=0}^{\infty} s(kT)\,\delta(\tau-kT)\Big]\,d\tau \qquad (13.92)$$

Thus we have obtained the convolution of two functions: $v_0(t)$ and $s_T(t)=\sum_{k=0}^{\infty} s(kT)\,\delta(t-kT)$. To the first corresponds the spectral density [see (2.67) and Fig. 2.15]

$$\mathbf{V}_0(\omega)=\tau_0\frac{\sin(\omega\tau_0/2)}{(\omega\tau_0/2)}e^{-i\omega\tau_0/2}$$

and to the function $\sum_{k=0}^{\infty} s(kT)\,\delta(t-kT)$ corresponds the spectral density [see (13.10)]

$$\mathbf{S}_T(\omega)=\frac{1}{T}\sum_{n=-\infty}^{\infty}\mathbf{S}\left(\omega-n\frac{2\pi}{T}\right)$$

Consequently, to time convolution (13.92) corresponds the spectral density equal to the product [see (2.64)]

$$\mathbf{S}'_T(\omega)=\left\{\tau_0\frac{\sin(\omega\tau_0/2)}{(\omega\tau_0/2)}e^{-i\omega\tau_0/2}\right\}\left\{\frac{1}{T}\sum_{n=-\infty}^{\infty}\mathbf{S}\left(\omega-n\frac{2\pi}{T}\right)\right\} \qquad (13.93)$$

The graph of the modulus of the function $\mathbf{S}'_T(\omega)$ for $\tau_0<T$ and $\tau_0=T$ is shown in Fig. 13.33. The dashed lines show the real spectra of the signal $s(t)$, which would be obtained with "slender" samples.

It is seen that the thickening of the pulses results in a deformation of the spectrum of the transmitted signal, this deformation being more pronounced for the higher frequencies of the signal.

The factor

$$K_{DAC}\left(\frac{\omega\tau_0}{2}\right)=\frac{\tau_0}{T}\,\frac{\sin(\omega\tau_0/2)}{\omega\tau_0/2} \tag{13.94}$$

in expression (13.93) can be regarded as the amplitude-frequency characteristic of the digital-to-analog converter (in Fig. 13.33, the function $K_{DAC}/(\tau_0/T)$ is shown by a dot-and-dash line).

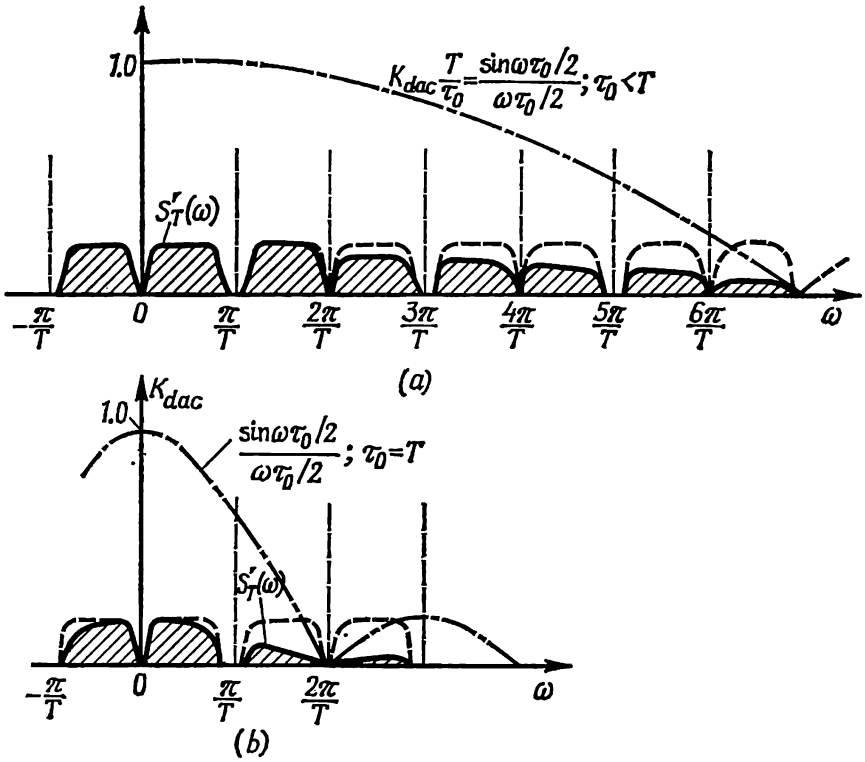


Fig. 13.33. Amplitude-frequency characteristic of a digital-to-analog converter and the spectral density of its output signal for (*a*) thin and (*b*) thick samples

In conclusion, let us consider the requirements for the amplitude-frequency characteristic $K_f(\omega)$ of the synthesizing filter. The ideal characteristic should have the form shown in Fig. 13.34 by a continuous line. If the spectrum of the useful signal is much narrower than the frequency range $(-f_1/2,\ f_1/2)$, the requirements for the slopes of the characteristic may be less rigid (see the dashed line in Fig. 13.34).

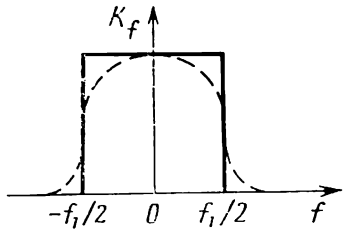


Fig. 13.34. Amplitude-frequency characteristic of an ideal (continuous line) and a real (dashed line) synthesizing filter

13.14. SPEED OF RESPONSE OF THE ARITHMETIC UNIT OF A DIGITAL FILTER. ROUNDING-OFF NOISE

Any digital filter circuit includes memory elements T, adders, and multipliers. A combination of these elements forms the arithmetic unit of the filter. (The switching devices required for synchronous recording and reading-out of binary symbols in the memory elements, and other auxiliary circuits are not considered here.)

The memory elements are a set of binary elements whose number is equal to the number r of digits.

The multipliers realizing the weighting coefficients a_0, a_1, a_2, . . . and b_1, b_2, . . . depend for their operation on the principle of digit-by-digit multiplication of all the digits of the input number by every digit of the number representing the given weighting coefficient, with subsequent addition of the partial products. The number $r_{a,\,b}$ of the binary digits, used for representing the weighting coefficient depends on the required calculation accuracy. In high-capacity computers, $r_{a,\,b}$ comes to 16 digits and more, while in digital filters, 4 to 6 digits are often adequate. To preserve the entire information contained in the input signal $s\,(t)$, the number of digits of the product of $s\,(t)$ and a_i or of $s\,(t)$ and b_i must be equal to the sum $r + r_{a,\,b}$. All the subsequent elements of the digital channel must be designed for this number of digits. The scope of necessary hardware can be reduced by rounding off the product (cutting off the lower-order digits). This results in an error called *rounding-off noise.*

The statistical properties of rounding-off noise are basically the same as those of quantization noise; the variance of rounding-off noise is taken at $\Delta^2_{a,\,b}/12$, where $\Delta_{a,\,b}$ is the step between the levels corresponding to the rejected digit of the product.

The speed of response is one of the most important characteristic of the arithmetic unit of a digital filter. This speed is defined by the number of operations that must be done during time T and the duration of one operation. The latter cannot be less than the operation time of the binary elements (flip-flops). The rapid and continuous development of microelectronics year after year cuts down the time lag of the electron devices used in computers. Nowadays the operation time of such devices amounts to a few nanoseconds.

Let us define the number of operations that must be performed during the time T when processing a signal in accordance with a given algorithm.

Let the convolution defined by expression (13.7) be the original algorithm. From this expression it is clear that n operations of multiplication and the same number of operations of addition are necessary for determining a single, nth sample of the output signal. With the number of samples in the processed realization of the signal being equal to $N \gg 1$, the total number of operations of multiplication is approximately equal to $(N/2)\, N = \frac{1}{2} N^2$ (the same number of operations of addition).

As mentioned above, the operation of multiplication is effected by multiple addition, the number of elementary addition operations being defined by the number of digits of the factors. With the duration of one operation of addition being equal to τ_1 and the number of digits equal to r, the total duration of processing of N samples is $T_{pr} = (N^2/2)\, r\tau_1$. Where real-time processing is required, T_{pr} must not exceed the duration $T_s = NT$ of the realization to be processed. From this stems the condition

$$(N^2/2)\, r\tau_1 \leqslant T_s = NT \text{ or } T \geqslant \frac{N}{2} r\tau_1$$

Substituting $T = 1/2f_m$ into this expression, we come to the following rough estimate of the highest possible frequency of the signal:

$$f_m \leqslant 1/Nr\tau_1$$

In particular, for $N = 1000$, $r = 10$, and $\tau_1 = 1$ ns

$$f_m \leqslant 1/10^3 \times 10 \times 10^{-9} = 10^5 \text{ Hz}$$

When processing shorter signals, e.g., with a base $N = 50$, the frequency can be brought up to 2 MHz.

We may see that at the present time the use of digital filters for *successive analysis* is limited to comparatively low frequencies.

In parallel-analysis systems with several channels, the speed of response can be considerably increased by using more expensive and complex equipment.

In principle, the speed of response can be brought close to τ_1, i.e., $f_m \leqslant 1/\tau_1$.

The main specific feature of a digital filter is the fact that its amplitude-frequency and phase-frequency characteristics are only determined by the weighting coefficients in the feedforward and feedback loops and the sampling interval T. This makes it possible to develop digital filters with such characteristics as are difficult to

realize, if ever, by means of ordinary inductive and capacitive filters.

High stability of frequency characteristics can be ensured by using quartz-crystal oscillators. Digital filters are reliable in operation, need no trimming, and are insensitive to ambient temperature and other operational conditions.

The simplicity of memory devices for storing digital information makes digital filters indispensable when processing signals requiring a time delay. Finally, it should be noted that digital filters are convenient for use in conjunction with computers.

Owing to these advantages, digital filters have found wide practical application in spite of their complex circuitry and the need to synchronize the operation of the electronic switches.

Chapter 14

REPRESENTATION OF OSCILLATIONS BY SOME SPECIAL FUNCTIONS

14.1. GENERAL

In Sec. 2.2 it was noted that various orthogonal sets of special, nonharmonic functions are required where it is necessary to approximate a given function $f(x)$ by a limited number of terms of a series. The conditions of orthonormality of these functions within a specified interval (a, b) are defined by

$$\int_a^b \varphi_n(x)\,\varphi_m(x)\,\rho(x)\,dx = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases} \tag{14.1}$$

This expression differs from definition (2.3) in that it contains the factor $\rho(x)$ under the integral sign, which is referred to as the *weighting* or *weight function*. It is said that the functions $\varphi_n(x)$ and $\varphi_m(x)$ are orthogonal with weight $\rho(x)$. This means that not these, but the functions $\sqrt{\rho(x)}\,\varphi_n(x)$ are orthogonal.

When determining the coefficients of the generalized Fourier series approximating the function $f(x)$, one should use a formula similar to (2.9), but with account being taken of the weighting function $\rho(x)$:

$$c_n = \frac{1}{\| \varphi_n \sqrt{\rho} \|^2} \int_a^b f(x)\,\varphi_n(x)\,\rho(x)\,dx \tag{14.2}$$

where

$$\| \varphi_n \sqrt{\rho} \|^2 = \int_a^b \varphi_n^2(x)\,\rho(x)\,dx \tag{14.3}$$

is the square of the norm of the function $\varphi_n(x)\sqrt{\rho(x)}$.

Continuous signals are most frequently represented by the Legendre, Chebyshev, Laguerre and Hermite orthogonal polynomials and functions which are described in Secs. 14.2 and 14.3. Discrete signals are represented by the Haar, Rademacher and Walsh step functions. The Walsh functions meeting the demands of computer engineering and microelectronics are of special importance. The continuous Walsh functions are considered in Secs. 14.4 to 14.6 and the discrete Walsh functions are studied in Sec. 14.7.

14.2. ORTHOGONAL POLYNOMIALS AND CONTINUOUS FUNCTIONS

Let us enumerate some of widely used polynomials and briefly discuss their properties.

1. The *Legendre polynomials* (*of the first kind*) defined by the formula

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n \tag{14.4}$$

are orthogonal with weight $\rho(x) = 1$ within the interval $-1 < x < 1$. For integral $n \geqslant 0$, the polynomials $P_n(x)$ contain a finite number of terms.

The lower-degree Legendre polynomials, shown graphically in Fig. 14.1, are defined by the expressions

$$P_0(x) = 1, \quad P_1(x) = x$$
$$P_2(x) = \frac{1}{2}(3x^2 - 1)$$
$$P_3(x) = \frac{1}{2}(5x^3 - 3x)$$
$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3) \tag{14.5}$$

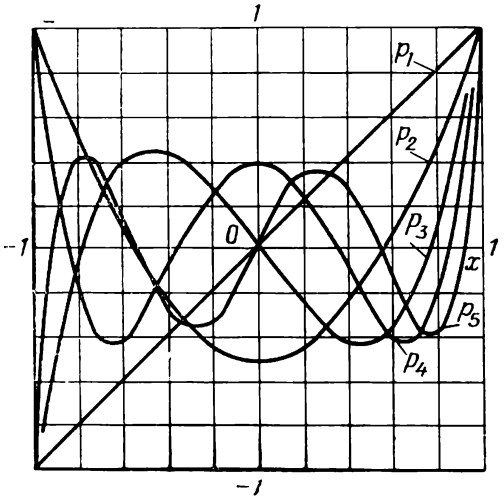


Fig. 14.1. Graphs of Legendre polynomials

In accordance with formula (14.4), the square of the norm of the function $P_n(x)$ is given by [15, 16]

$$\| P_n(x) \|^2 = \int_{-1}^{1} P_n^2(x)\, dx = \frac{2}{2n+1} \tag{14.6}$$

In this case, expression (2.9) for the coefficients c_n assumes the form

$$c_n = \frac{2n+1}{2} \int_{-1}^{1} f(x) P_n(x)\, dx \tag{14.7}$$

and series (2.8) is written as

$$f(x) = c_0 P_0(x) + c_1 P_1(x) + \ldots + C_n P_n(x) + \ldots \tag{14.8}$$

2. The *Chebyshev polynomials* (*of the first kind*) are defined as

$$T_n(x) = \frac{(-2)^n n!}{(2n)!} \sqrt{1 - x^2} \frac{d^n}{dx^n} \left(\sqrt{1 - x^2}\right)^{2n-1} \tag{14.9}$$

The lower-degree Chebyshev polynomials are

$$T_0(x) = 1,\ T_1(x) = x,\ T_2(x) = 2x^2 - 1$$
$$T_3(x) = 4x^3 - 3x,\ T_4(x) = 8x^4 - 8x^2 + 1$$
$$T_5(x) = 16x^5 - 20x^3 + 5x$$

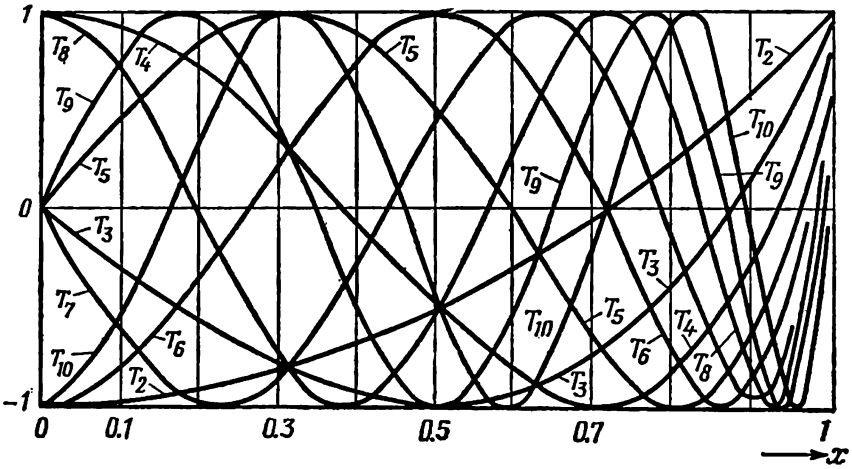


Fig. 14.2. Graphs of Chebyshev polynomials

Figure 14.2 shows the graphs of the polynomials $T_n(x)$ within the interval $-1 < x < 1$, and Fig. 14.3 shows the graph of one of them, in particular, of the fourth-degree polynomial, for $0 < |x| < 5/4$. For $|x| > 1$, $T_n(x)$ tends to infinity as $2^{n-1}x^n$.

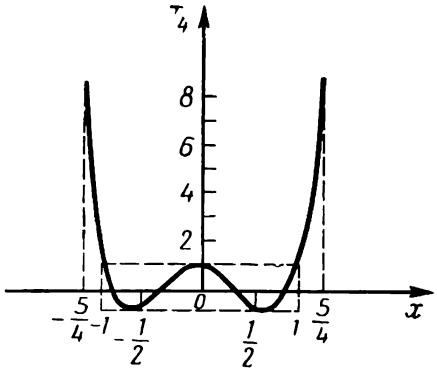


Fig. 14.3. Graphs of fourth-degree Chebyshev polynomials

An important feature of the Chebyshev polynomials is that they, among all polynomials of degree n and with the leading coefficient equal to unity, show the *least tendency to deviate from zero within the interval* $-1 < x < 1$. Owing to this property, the Chebyshev polynomials ensure the least maximum error of uniform approximation within the interval $-1 < x < 1$.

The Chebyshev polynomials are not orthogonal; however, after multiplication by $1/\sqrt[4]{1-x^2}$, they form an orthogonal (within the interval $-1 < x < 1$) set of functions $(1/\sqrt[4]{1-x^2})\,T_n(x)$. In other words, the polynomials $T_n(x)$ are orthogonal with weight $\rho(x) = 1/\sqrt{1-x^2}$

$$\int_{-1}^{1} T_n(x)\,T_m(x)\,\frac{dx}{\sqrt{1-x^2}} = \begin{cases} 0 & \text{if } n \neq m \\ \frac{\pi}{2} & \text{if } n = m \end{cases} \qquad (14.10)$$

Furthermore, if $m = n = 0$

$$\int_{-1}^{1} T_n(x)\, T_m(x) \frac{dx}{\sqrt{1-x^2}} = \int_{-1}^{1} \frac{dx}{\sqrt{1-x^2}} = \pi \qquad (14.10')$$

Thus, the norm $\| T_0\sqrt{\rho} \| = \sqrt{\pi}$ and $\| T_n\sqrt{\rho} \| = \sqrt{\pi/2}$.

When expanding the function $f(x)$ in the Chebyshev polynomials (taking into account that $T_0(x) = 1$), the coefficients of the series

$$f(x) = c_0 + \sum_{n=1}^{\infty} c_n T_n(x), \quad -1 < x < 1$$

must be determined, in accordance with (14.2), (14.10) and (14.10'), by the following expressions:

$$c_0 = \frac{1}{\pi} \int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}\, dx, \quad c_n = \frac{2}{\pi} \int_{-1}^{1} \frac{f(x)\, T_n(x)}{\sqrt{1-x^2}}\, dx \qquad (14.11)$$

The behaviour of the Chebyshev polynomials within the interval $-1 < x < 1$, in conjunction with the infinite increase of $| T_n(x) |$ for $| x | > 1$, makes these polynomials very effective for approximation of the amplitude-frequency characteristics of different filters. This question is discussed in Chapter 15.

3. The *Laguerre polynomials* are defined by the formula

$$L_n(x) = \frac{e^x d^n}{n!\, dx^n} (x^n e^{-x}), \quad x \geqslant 0 \qquad (14.12)$$

The first four polynomials are

$$L_0(x) = 1, \; L_1(x) = -x + 1, \; L_2(x) = x^2/2 - 2x + 1$$
$$L_3(x) = -x^3/6 + 3x^2/2 - 3x + 1$$

The Laguerre polynomials are orthogonal on the semi-axis $0 < x < \infty$ with weight $\rho(x) = e^{-x}$.

Since the Laguerre polynomials form a set of functions that diverge as $x \to \infty$, it is more convenient to use the *Laguerre functions*

$$l_n(x) = \sqrt{\rho(x)}\, L_n(x) = e^{-x/2} L_n(x) \qquad (14.13)$$

The Laguerre functions $l_n(x)$ are orthogonal with unit weight. Figure 14.4 illustrates the Laguerre functions for $n = 1, 2, \ldots, 5$. The norm of the function $l_n(x)$ is

$$\| l_n \| = \sqrt{\int_0^{\infty} l_n^2(x)\, dx} = 1$$

therefore, when expanding the function $f(x)$ in the Laguerre functions, the coefficients of the series

$$f(x) = \sum_{n=0}^{\infty} c_n l_n(x) \qquad (14.14)$$

must be defined by the formula

$$c_n = \int_0^\infty f(x)\, l_n(x)\, dx \tag{14.15}$$

The Laguerre functions have found wide application in measuring technique and in multichannel communication systems, which, in a great measure, is explained by the simplicity of their generation. The thing is that the function $l_n(t)$ coincides in shape with the im-

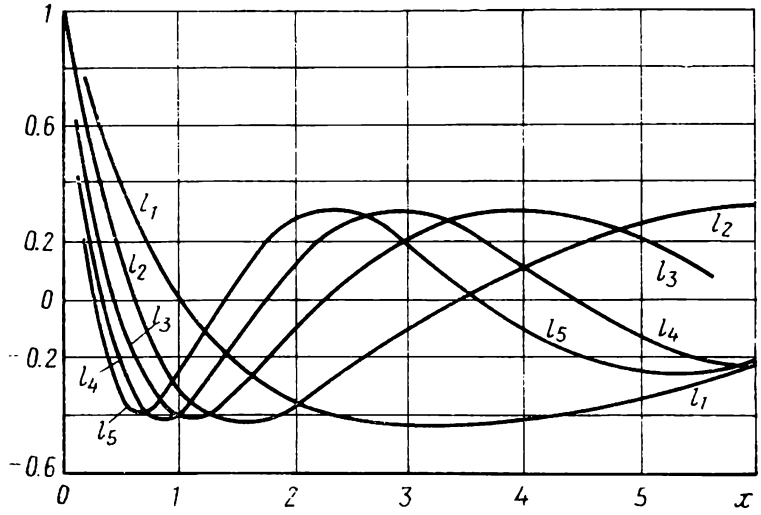


Fig. 14.4. Laguerre functions

pulse response of a physical circuit consisting of simple links connected in cascade (Fig. 14.5). To determine the transfer function of this circuit, let us apply the Laplace transformation to the Laguerre function (14.13), having preliminarily replaced in (14.12) and (14.13) the variable x by a new variable $x = \alpha t$:

$$l_n(\alpha t) = \frac{e^{\alpha t/2}}{n!} \frac{d^n}{dt^n} (t^n e^{-\alpha t})$$

To the time function $e^{-\alpha t} t^n$ corresponds the transform $n!/(p+\alpha)^{n+1}$, and to n-multiple differentiation corresponds the multiplication of the transform by p^n. Also, taking into account the fact that the multiplication by $e^{\alpha t/2}$ gives a shift on the p-plane by an amount $-\alpha/2$, we obtain the following transform for the Laguerre function:

$$\frac{(p-\alpha/2)^n}{(p+\alpha/2)^{n+1}} = \frac{1}{(p+\alpha/2)} \left(\frac{p-\alpha/2}{p+\alpha/2} \right)^n$$

The transfer function $1/(p+\alpha/2)$ of the first link is realized by an integrating R-C circuit satisfying the condition $RC = 2/\alpha$. The transfer function $(p-\alpha/2)/(p+\alpha/2)$ corresponds to a bridge circuit with $RC = 2/\alpha$.

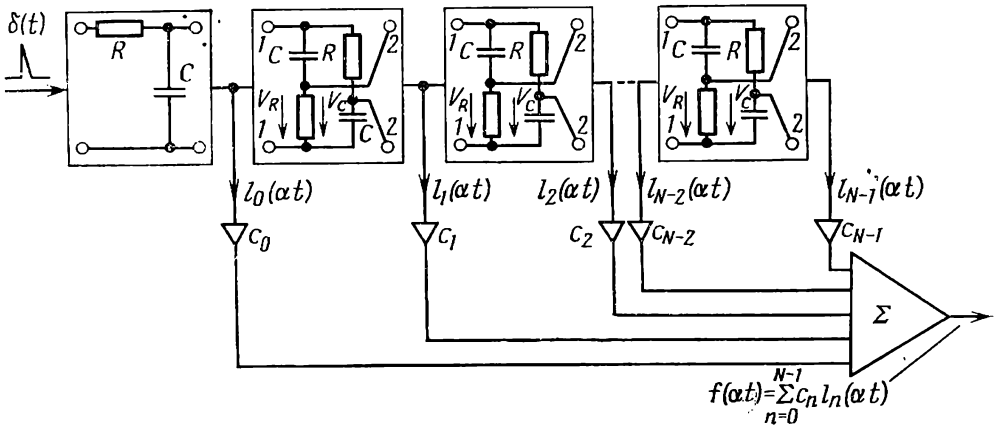


Fig. 14.5. Laguerre function generator

In fact, directly from the bridge circuit of one link (Fig. 14.5) follow the relations

$$\mathbf{V}_{2\text{-}2}/\mathbf{V}_{1\text{-}1} = (\mathbf{V}_R - \mathbf{V}_C)/\mathbf{V}_{1\text{-}1} = (R - 1/i\omega C)/(R + 1/i\omega C)$$

whence

$$\mathbf{V}_{2\text{-}2}(p)/\mathbf{V}_{1\text{-}1}(p) = (p - 1/RC)/(p + 1/RC)$$

When the circuit (Fig. 14.5) is excited by a delta function, the output oscillation of the first link will be $e^{-\alpha t/2} = l_0(\alpha t)$, and the oscillations at the outputs of the next links will respectively be $l_1(\alpha t)$, $l_2(\alpha t)$, etc.

The weighted addition of all these oscillations yields the following oscillation at the adder output:

$$f(\alpha t) = \sum_{n=0}^{N-1} C_n l_n(\alpha t), \quad t \geqslant 0$$

4. The *Hermite polynomials* are defined by the formula

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}(e^{-x^2}) \tag{14.16}$$

The first five Hermite polynomials are

$$H_0(x) = 1, \; H_1(x) = 2x, \; H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x, \; H_4(x) = 16x^4 - 48x^2 + 12$$

The graphs of these polynomials are shown in Fig. 14.6.

The Hermite polynomials are orthogonal with weight $\rho(x) = e^{-x^2}$ on the entire axis $-\infty < x < \infty$, so that

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) e^{-x^2} dx = \begin{cases} 0 & \text{if } m \neq n \\ 2^n \sqrt{\pi}\, n! & \text{if } m = n \end{cases}$$

Thus, the norm of the function $H_n(x)\sqrt{\rho(x)} = H_n(x)\, e^{-x^2/2}$ is equal to

$$\| H_n \sqrt{\rho} \| = \sqrt{2^n \sqrt{\pi}\, n!}$$

To transform to the orthonormalized set of Hermite polynomials, one introduces the function

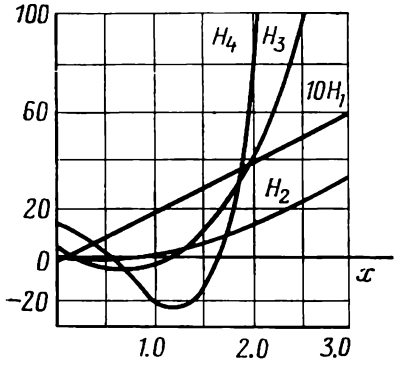


Fig. 14.6. Graphs of Hermite polynomials

$$\varphi_n(x) = \frac{H_n(x)\sqrt{\rho(x)}}{\| H_n \sqrt{\rho} \|} = \frac{H_n(x)\, e^{-x^2/2}}{\sqrt{2^n \sqrt{\pi}\, n!}} \qquad (14.17)$$

In this case, the expansion of the function $f(x)$ in the normalized Hermite functions is written in the form

$$f(x) = \sum_{n=0}^{\infty} c_n \varphi_n(x)$$

where

$$c_n = \int_{-\infty}^{\infty} f(x)\, \varphi_n(x)\, dx \qquad (14.18)$$

The graphs of the normalized functions $\varphi_n(x)$ are shown in Fig. 14.7.

Thus, one may subdivide the orthogonal sets of functions into two classes: (1) sets defined within a finite interval (Legendre and Chebyshev polynomials) and (2) sets defined within an infinite interval which is either the semi-axis $0 < x < \infty$ (Laguerre polynomials) or of the entire axis $-\infty < x < \infty$ (Hermite polynomials). The orthogonal sets of the first kind are to be used for approximation of the processes and characteristics defined within a finite interval, while the sets of the second kind are preferably used for the functions $f(x)$ specified within an infinite interval.

The type of the weighting function $\rho(x)$ is very important for the choice of polynomials: the weighting function $\rho(x)$ must be maximum in the interval where the best approximation is needed. In this case it is possible to reduce the number of terms of the series with the given admissible approximation error. Selecting an appropriate weighting function also makes it possible to approximate the processes of finite

duration by the polynomials of the second kind (defined within an infinite interval). For this it is necessary that the effective duration

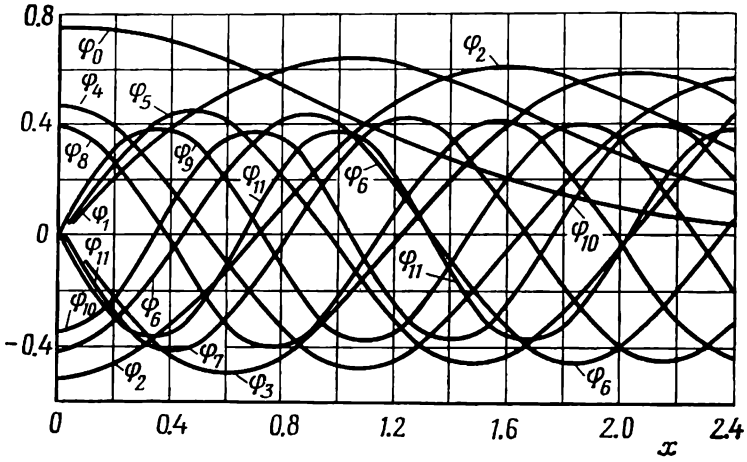


Fig. 14.7. Graphs of normalized Hermite functions

of the weighting function be close to the duration of the signal being approximated.

14.3. EXAMPLES OF APPLICATION OF CONTINUOUS FUNCTIONS

To simplify calculations and to make the analysis more graphic, the functions to be approximated will be given in an analytic form. In this connection, it should be kept in mind that under actual conditions it is often necessary to approximate functions obtained experimentally and given in the form of graphs, tables, oscillograms, etc.

Let us consider the expansion of a half-wave cosine pulse (Fig. 14.8) in different orthogonal polynomials. Figure 14.8 illustrates the function

$$s(t) = \cos(\pi t/\tau_p), \quad -\tau_p/2 < t < \tau_p/2 \qquad (14.19)$$

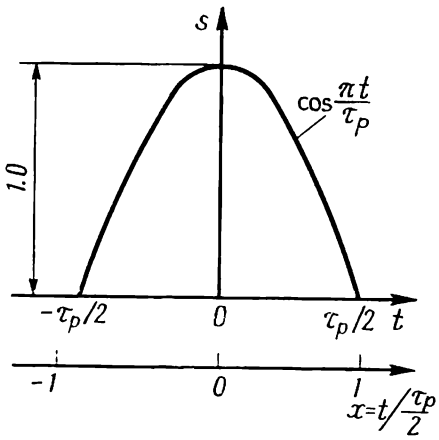


Fig. 14.8. Half-wave cosine pulse

Transforming to the dimensionless time $x = \frac{t}{\tau_p/2}$, we obtain

$$s(t) = s\left(\frac{\tau_p}{2}x\right) = \cos\left(\frac{\pi}{2}x\right), \quad -1 < x < 1 \qquad (14.20)$$

Since the function $s\left(\frac{\tau_p}{2}x\right)$ is specified within a finite interval $(-1, 1)$, it is expedient to analyze the expansion of this function in the Legendre and Chebyshev polynomials. However, let us preliminarily give the expansion of the function $s(t)$ in trigonometric functions with a period T equal to the duration τ_p of the pulse, corresponding to the dimensionless period $x_T = 2$ (Fig. 14.8).

Since $s\left(\frac{\tau_p}{2}x\right)$ is an even function, we obtain a cosine series

$$s\left(\frac{\tau_p}{2}x\right) = \frac{a_0}{2} + a_1 \cos\left(\frac{2\pi}{x_T}x\right) + a_2 \cos\left(2\frac{2\pi}{x_T}x\right) + \ldots$$

$$= \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\pi x), \quad -1 < x < 1 \qquad (14.21)$$

whose coefficients, in accordance with (2.32), are

$$\frac{a_0}{2} = \frac{1}{x_T}\int_{-1}^{1} \cos\left(\frac{\pi}{2}x\right) dx = \frac{2}{x_T}\frac{\sin \pi/2}{\pi/2} = \frac{2}{\pi}$$

$$a_n = \frac{2}{x_T}\int_{-1}^{1} \cos\left(\frac{\pi}{2}x\right)\cos(n\pi x)\, dx$$

$$= \frac{4}{x_T}\int_{0}^{1} \frac{1}{2}\left[\cos\left(n-\frac{1}{2}\right)\pi x + \cos\left(n+\frac{1}{2}\right)\pi x\right] dx$$

$$= \frac{2}{\pi}\left[\frac{1}{2n-1}\sin\left(n-\frac{1}{2}\right)\pi + \frac{1}{2n+1}\sin\left(n+\frac{1}{2}\right)\pi\right]$$

By direct calculations we find

$$a_1 = \frac{2}{\pi}\frac{2}{3}, \quad a_2 = -\frac{2}{\pi}\frac{2}{15}, \quad a_3 = \frac{2}{\pi}\frac{2}{35}, \quad a_4 = -\frac{2}{\pi}\frac{2}{63}$$

The spectrum of the function $s\left(\frac{\tau_p}{2}x\right)$ in the basic set of trigonometric functions $\cos n\pi x$ is shown in Fig. 14.9*a*, while the graphs of the individual terms of series (14.21) are shown in Fig. 14.9*b*.

When using the Legendre polynomials, the coefficients of series (14.8), in accordance with (14.7), will be

$$c_n = \frac{2n+1}{2}\int_{-1}^{1} \cos\left(\frac{\pi}{2}x\right) P_n(x)\, dx \qquad (14.22)$$

Thus,

$$c_0 = \frac{1}{2}\int_{-1}^{1} \cos\left(\frac{\pi}{2}x\right) dx = \frac{2}{\pi}$$

$$c_1 = \frac{2+1}{2}\int_{-1}^{1} \cos\left(\frac{\pi}{2}x\right) x\,dx = 0 \text{ (integral of odd function)}$$

$$c_2 = \frac{2\times 2+1}{2}\int_{-1}^{1} \cos\left(\frac{\pi}{2}x\right)\left[\frac{1}{2}(3x^2-1)\right] dx = -0.68$$

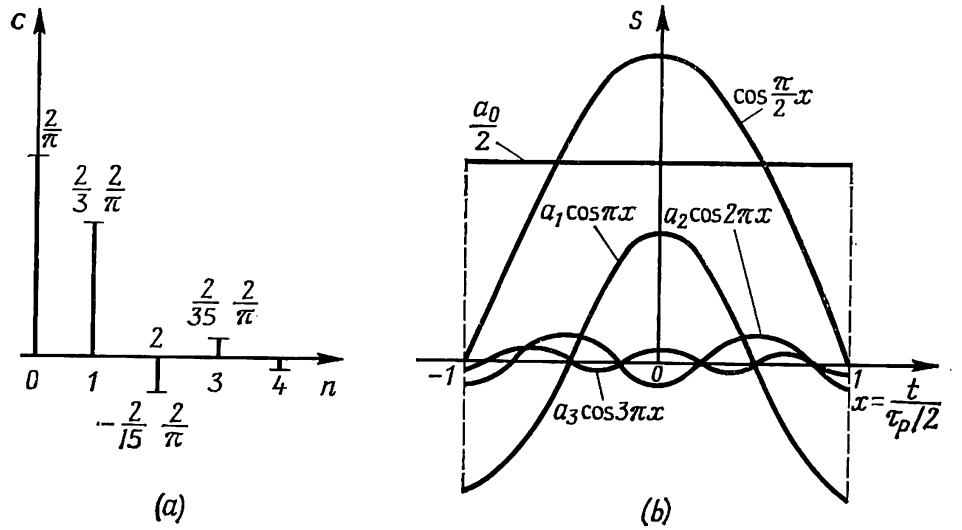


Fig. 14.9. (*a*) Spectrum of a half-wave cosine pulse in the basic set of trigonometric functions and (*b*) the first four terms of the Fourier series

$c_3 = 0$ (integral of odd function)

$$c_4 = \frac{2\times 4+1}{2}\int_{-1}^{1} \cos\left(\frac{\pi}{2}x\right)\left[\frac{1}{8}(35x^4-30x^2+3)\right] dx = -0.03$$

Thus,

$$\cos\left(\frac{\pi}{2}x\right) = \frac{2}{\pi} - 0.68P_2(x) - 0.03P_4(x) + \ldots, \quad -1 < x < 1$$

Using tables of the Legendre polynomials [15], it is possible to plot the graph of $s\left(\frac{\tau_p}{2}x\right)$, shown in Fig. 14.10. In practice, two or three terms of the series are sufficient for satisfactory approximation of the function

$$s(t) = \cos\left(\frac{\pi}{2}x\right), \quad -\frac{\tau_p}{2} < t < \frac{\tau_p}{2}, \quad -1 < x < 1$$

Now let us consider the expansion of this function in the Chebyshev polynomials. The calculation of the coefficients c_n for the function $\cos\left(\frac{\pi}{2}x\right)$ is facilitated by the formula {see [16], formula (7.355.2)}

$$\int_0^1 T_{2k}(x)\cos ax\,\frac{dx}{\sqrt{1-x^2}} = (-1)^k\frac{\pi}{2}J_{2k}(a) \qquad (14.23)$$

in which $J_{2k}(a)$ is a Bessel function. In this case, formulas (14.11)

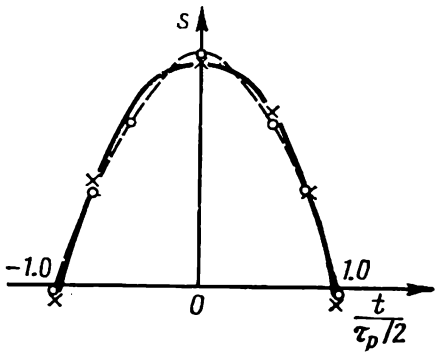


Fig. 14.10. Approximation of a half-wave cosine pulse by the first three Legendre polynomials (continuous line) and Chebyshev polynomials (dashed line)

assume the form

$$c_0 = \frac{1}{\pi}\int_{-1}^{1}\cos\left(\frac{\pi}{2}x\right)T_0(x)\,\frac{dx}{\sqrt{1-x^2}} = \frac{2}{\pi}\frac{\pi}{2}J_0(\pi/2) = J_0(\pi/2)$$

$$c_2 = \frac{2}{\pi}\int_{-1}^{1}\cos\left(\frac{\pi}{2}x\right)T_2(x)\,\frac{dx}{\sqrt{1-x^2}} = -2J_2(\pi/2)$$

$$c_4 = \frac{2}{\pi}\int_{-1}^{1}\cos\left(\frac{\pi}{2}x\right)T_4(x)\,\frac{dx}{\sqrt{1-x^2}} = 2J_4\left(\frac{\pi}{2}\right),\ \text{etc.}$$

The functions $J_0(m)$ and $J_1(m)$ are tabulated [15], while for higher orders $n = 2, 3, 4, \ldots$ one can use the relations

$$J_{n+1}(m) = \frac{2n}{m}J_n(m) - J_{n-1}(m) \qquad (14.24)$$

For the argument $m = a = \pi/2$, we find from the tables $J_0(\pi/2) = 0.473$ and $J_1(\pi/2) = 0.566$. Then

$$J_2(\pi/2) = \frac{2}{\pi/2}J_1(\pi/2) - J_0(\pi/2) = \frac{2}{\pi/2}0.566 - 0.473 = 0.25$$

Further

$$J_3\left(\frac{\pi}{2}\right)=\frac{2\times 2}{\pi/2}J_2\left(\frac{\pi}{2}\right)-J_1\left(\frac{\pi}{2}\right)=0.071$$

$$J_4\left(\frac{\pi}{2}\right)=\frac{2\times 3}{\pi/2}J_3\left(\frac{\pi}{2}\right)-J_2\left(\frac{\pi}{2}\right)=0.02$$

Thus, $c_0 = 0.473$, $c_2 = -0.5$, $c_4 = 0.04$, etc. Finally, we have

$$\cos\left(\frac{\pi}{2}x\right)=0.473-0.5T_2(x)$$
$$+0.04T_4(x)-\ldots,\quad -1<x<1 \qquad (14.25)$$

The sum of the first three terms of the series is shown in Fig. 14.10 (dashed line). It is evident that for satisfactory approximation of the function $\cos\left(\frac{\pi}{2}x\right)$ three terms of the series are sufficient.

From comparison between the three approximations of the function $\cos\left(\frac{\pi}{2}x\right)$ it is clear that the trigonometric Fourier series requires addition of a greater number of the expansion terms than the Legendre or Chebyshev polynomials. Of the latter two expansions, the one in the Legendre polynomials gives a lower error at the point $x = 0$ (when only two terms of the series are added together.) However, at the point $x = 1$ the approximation to the exact value (zero) is better by Chebyshev than by Legendre. This is explained by the fact that the weighting function $\rho(x) = 1/\sqrt{1-x^2}$ increases as x approaches unity.

From this comparison one should not conclude that the representation of functions in the form of trigonometric series is in all cases less preferable than expansion in polynomials.

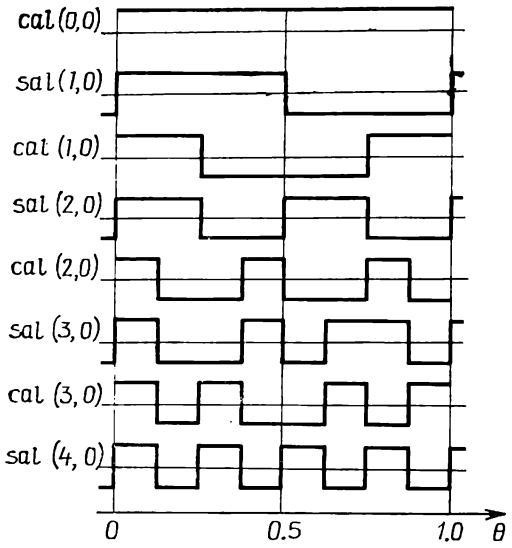
14.4. DEFINITION OF THE WALSH FUNCTIONS

The Walsh and Rademacher functions, known since 1922, were not applied for a long time. New interest in these functions and their wide application at the present time are associated with the development of computer engineering.

There are different methods of defining the Walsh functions. Let us consider the method based on the interrelation between the Walsh and Rademacher functions. The Rademacher functions are obtained from sinusoidal functions with the aid of the relation

$$r_k(\theta) = \operatorname{sgn}[\sin(2^k\pi\theta)],\quad 0 \leqslant \theta < 1 \qquad (14.26)$$

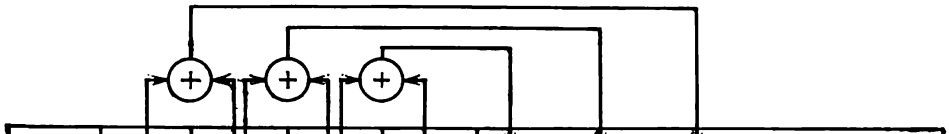
where the argument $\theta = t/T_0$ is dimensionless time, i.e., time normalized with reference to an arbitrary interval T_0, and the positive integer k is the order of the function. The symbol "sgn" (sign func-



wal(w,θ)	pal(p,θ)	had(h,θ)
w	p	h
0	0	0
1	1	4
2	3	6
3	2	2
4	6	3
5	7	7
6	5	5
7	4	1

Fig. 14.12. First eight Walsh functions and their numbering with different methods of ordering

Table 14.1



w	w_0	w_1	w_2	w_3	$r_1(\theta) \times r_2(\theta) \times r_3(\theta) = \mathrm{wal}(w,\theta)$
0	0	0	0	0	$r_1^0(\theta) \times r_2^0(\theta) \times r_3^0(\theta) = \mathrm{wal}(0,\theta)$
1	0	0	0	1	$r_1^1(\theta) \times r_2^0(\theta) \times r_3^0(\theta) = \mathrm{wal}(1,\theta)$
2	0	0	1	0	$r_1^1(\theta) \times r_2^1(\theta) \times r_3^0(\theta) = \mathrm{wal}(2,\theta)$
3	0	0	1	1	$r_1^0(\theta) \times r_2^1(\theta) \times r_3^0(\theta) = \mathrm{wal}(3,\theta)$
4	0	1	0	0	$r_1^0(\theta) \times r_2^1(\theta) \times r_3^1(\theta) = \mathrm{wal}(4,\theta)$
5	0	1	0	1	$r_1^1(\theta) \times r_2^1(\theta) \times r_3^1(\theta) = \mathrm{wal}(5,\theta)$
6	0	1	1	0	$r_1^1(\theta) \times r_2^0(\theta) \times r_3^1(\theta) = \mathrm{wal}(6,\theta)$
7	0	1	1	1	$r_1^0(\theta) \times r_2^0(\theta) \times r_3^1(\theta) = \mathrm{wal}(7,\theta)$

In this table use is made of the following designations: w is the serial number of a function in the set, w_m is the mth digit of the number w in the binary system, i.e.,

$$w=(w_1w_2\ldots w_m\ldots w_n)_2=w_12^{n-1}+w_22^{n-2}+\ldots+w_m2^{n-m}$$

$$+\ldots\ldots+w_n2^0=\sum_{m=1}^{n}w_m2^{n-m}=\sum_{m=1}^{n}w_{n-m+1}2^{m-1}$$

$$w_m=0;\quad 1,\ w_0=0 \tag{14.28}$$

and $\oplus$ is the symbol for digit-by-digit modulo 2 addition according to the rules

$$1\oplus 1=0\oplus 0=0,\ 1\oplus 0=0\oplus 1=1 \tag{14.29}$$

The method of forming the Walsh functions, shown in Table 14.1, can be expressed analytically for any $N=2^n$ in the form of the following relation:

$$\text{wal}\,(w,\ \theta)=\prod_{k=1}^{n}[r_k(\theta)]^{w_{n-k+1}\oplus w_{n-k}} \tag{14.30}$$

Let us illustrate the application of expression (14.30) with an example of the sixth ($w=6$) Walsh function entering into a set of $N=2^3=8$ functions. The product in (14.30) consists of three factors (for $k=1, 2$, and 3) of the form $[r_1(\theta)]^{w_3\oplus w_2}$, $[r_2(\theta)]^{w_2\oplus w_1}$ and $[r_3(\theta)]^{w_1\oplus w_0}$. Substituting $w=6$ and $n=3$ into the left-hand side of (14.28), we get $6=w_12^2+w_22^1+w_32^6$, whence $w_1=1$, $w_2=1$, $w_3=0$. Thus,

$$w_3\oplus w_2=0\oplus 1=1$$
$$w_2\oplus w_1=1\oplus 1=0$$
$$w_1\oplus w_0=1\oplus 0=1$$

and by formula (14.30)

$$\text{wal}\,(6,\ \theta)=r_1(\theta)\,r_2^0(\theta)\,r_3(\theta)=r_1(\theta)\,r_3(\theta)$$

From Fig. 14.12 it is evident that to even w correspond the functions wal (w, θ) which are even with respect to the middle of the interval of definition ($\theta=0.5$), while to odd w correspond the odd functions. Such a reciprocal one-to-one correspondence between the parity of the functions wal (w, θ) and the parity of their numbers w is similar to the properties of the trigonometric functions

$$\cos\left(k\frac{2\pi}{T}t\right)\text{ and }\sin\left(k\frac{2\pi}{T}t\right)\text{ (see Fig. 14.13).}$$

Therefore, the symbols cal (j, θ) for even Walsh functions and sal (j, θ) for odd Walsh functions are sometimes used.

One can easily verify that the functions cal (j, θ) and sal (j, θ) are related to the functions wal (w, θ) by the following relations:

$$\text{cal}\,(j, \theta) = \text{wal}\,(2j, \theta),\ \text{sal}\,(j, \theta) = \text{wal}\,(2j - 1, \theta)$$

These designations are shown in Fig. 14.12.

The method of numbering of the functions in a set is called *ordering*. The Walsh functions formed by means of expression 14.30 are ordered by Walsh.

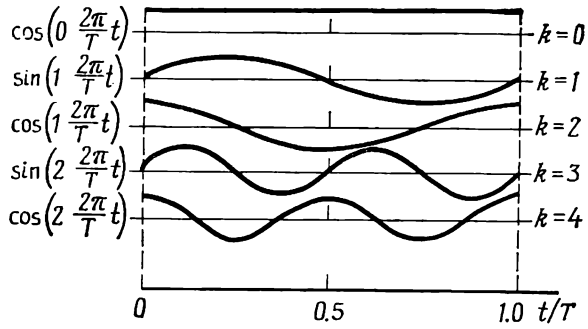


Fig. 14.13. Even parity of the numbers of cosine functions and the odd parity of the numbers of sine functions

Use is often made of Walsh functions ordered by Paley [pal (p, θ)] and by Hadamard [had (h,θ)].

No matter what the ordering, the Walsh functions constituting a set of $N = 2^n$ functions can always be represented in the form of the product of the like powers of the first n Rademacher functions, while the principle of finding the exponents is individual for each ordering.

Thus, for the ordering by Paley, the principle of finding these powers is illustrated in Table 14.2 by the example of $N = 2^3 = 8$. In this table, by analogy with (14.28), the serial number p of the functions pal (p, θ) is given in the binary form

$$p = (p_1 p_2 \ldots p_m \ldots p_n)_2 = \sum_{m=1}^{n} p_{n-m+1} 2^{m-1} \qquad (14.28')$$

It is clear that analytically the Walsh functions ordered by Paley are written in the form

$$\text{pal}\,(p,\theta) = \prod_{k=1}^{n} [r_k(\theta)]^{p_{n-k+1}} \qquad (14.31)$$

Comparing the methods of forming the exponents of the Rademacher functions, given in the examples of Tables 14.1 and 14.2, one can easily come to the conclusion that the binary digits of the serial numbers of the Walsh functions ordered by Paley are related to the

Table 14.2

p	p_1	p_2	p_3	$r_1(\theta) \times r_2(\theta) \times r_3(\theta) = \text{pal}(p,\theta)$
0	0	0	0	$r_1^0(\theta) \times r_2^0(\theta) \times r_3^0(\theta) = \text{pal}(0,\theta)$
1	0	0	1	$r_1^1(\theta) \times r_2^0(\theta) \times r_3^0(\theta) = \text{pal}(1,\theta)$
2	0	1	0	$r_1^0(\theta) \times r_2^1(\theta) \times r_3^0(\theta) = \text{pal}(2,\theta)$
3	0	1	1	$r_1^1(\theta) \times r_2^1(\theta) \times r_3^0(\theta) = \text{pal}(3,\theta)$
4	1	0	0	$r_1^0(\theta) \times r_2^0(\theta) \times r_3^1(\theta) = \text{pal}(4,\theta)$
5	1	0	1	$r_1^1(\theta) \times r_2^0(\theta) \times r_3^1(\theta) = \text{pal}(5,\theta)$
6	1	1	0	$r_1^0(\theta) \times r_2^1(\theta) \times r_3^1(\theta) = \text{pal}(6,\theta)$
7	1	1	1	$r_1^1(\theta) \times r_2^1(\theta) \times r_3^1(\theta) = \text{pal}(7,\theta)$

binary digits of the serial numbers of the Walsh functions ordered by Walsh as follows:

$$p_m = w_{m-1} \oplus w_m \tag{14.32}$$

Thus, the transformation from the ordering by Walsh to the ordering by Paley consists in transposing the functions in the set in accord with (14.32).

The functions had (h, θ) can be formed by using the Hadamard matrices. The Hadamard matrix H_N of order $N = 2^n$ is a square $N \times N$ matrix with elements ± 1, such that

$$H_N \times H_N^T = NI$$

where I is the unit matrix and T is the transpose symbol.

The normalized Hadamard matrix of order N can be constructed recursively, i.e.,

$$H_N = \begin{bmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{bmatrix} \quad \text{if} \quad H_1 = 1 \tag{14.33}$$

Thus, for example,

$$H_2 = \begin{bmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$H_4 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} = \left[\begin{array}{cc:cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hdashline 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{array}\right], \quad \text{etc.}$$

The Hadamard-ordered Walsh function, i.e., had (h, θ), with the serial number h is a train of square pulses with unit amplitudes and polarities corresponding to the signs of the elements of the hth line of the Hadamard matrix. By the pulse duration is understood the $(1/N)$th fraction of the interval [0, 1].

To illustrate the relation between the function had (h, θ) and the Hadamard matrix and to determine the position of these functions in the set, let us give the Hadamard matrix for $N = 8 = 2^3$, replacing 1 and —1 by the plus and the minus sign, respectively:

$$H_8 = \begin{bmatrix} H_4 & H_4 \\ H_4 & -H_4 \end{bmatrix} = \left[\begin{array}{cccc:cccc} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ \hdashline + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{array}\right] \begin{matrix} h_0 \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \end{matrix}$$

The relation between the ordering by Hadamard and that by Paley is defined by the expression

$$h_k = p_{n-k+1} \tag{14.34}$$

where $n = \log_2 N$.

The numbering of the first eight Walsh functions by different ordering methods is given in the Table of Fig. 14.12.

It should be noted that the above orderings stem from the symmetry of the Hadamard matrix, consisting in that the transposed matrix coincides with the original one: $H_N = H_N^T$. As seen from the preceding text, the above orderings correspond to the symmetry of the associated matrices.

One should not think that the orderings by Walsh, Paley, and Hadamard are the only ones possible.

Regardless of the method of ordering, the Walsh functions will further be designated by the symbol wal (i, θ). The Walsh functions are *orthonormalized* within the interval $0 \leqslant \theta \leqslant 1$:

$$\int_0^1 \text{wal}\ (k, \theta)\ \text{wal}\ (i, \theta)\, d\theta = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases} \tag{14.35}$$

The Walsh functions are *multiplicative*, i.e., the multiplication of two Walsh functions yields another Walsh function, so that

$$\text{wal}\ (k,\ \theta)\ \text{wal}\ (i,\ \theta) = \text{wal}\ (k \oplus i,\ \theta) \qquad (14.36)$$

The Walsh functions wal (i, θ) have the property of *symmetry* so that all conclusions as to i are valid for θ. For example, multiplicative property (14.36), with account being taken of the symmetry property, will be written as

$$\text{wal}\ (i,\ \theta_1)\ \text{wal}\ (i,\ \theta_2) = \text{wal}\ (i,\ \theta_1 \oplus \theta_2) \qquad (14.37)$$

Multiplication of any Walsh function by itself yields a zero-order function wal $(0, \theta)$, since in this case we have only such products as $(+1)(+1)$ and $(-1)(-1)$. Thus,

$$\text{wal}\ (i,\theta)\ \text{wal}(i,\theta) = \text{wal}(0,\theta)$$

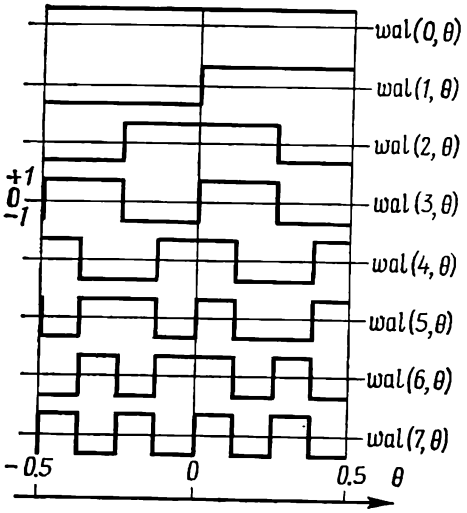


Fig. 14.14. First eight Walsh functions within the interval $-0.5 \leqslant \theta < 0.5$.

It is also obvious that multiplication of wal (i, θ) by wal $(0, \theta)$ does not change the function wal (i, θ).

The Walsh functions are sometimes defined within the interval $-1/2 \leqslant \theta < 1/2$. The first eight functions within this interval are shown in Fig. 14.14. The Walsh functions can serve as a basis for a spectral (nonharmonic) representation of signals.

Any function $f(\theta)$ integrable over the interval $0 \leqslant \theta < 1$ can be represented by a Fourier series in a set of Walsh functions

$$f(\theta) = A(0) + A(1)\,\text{wal}\,(1, \theta) + A(2)\,\text{wal}\ (2,\ \theta) + \ldots + A(i)\,\text{wal}\ (i,\ \theta) + \ldots \qquad (14.38)$$

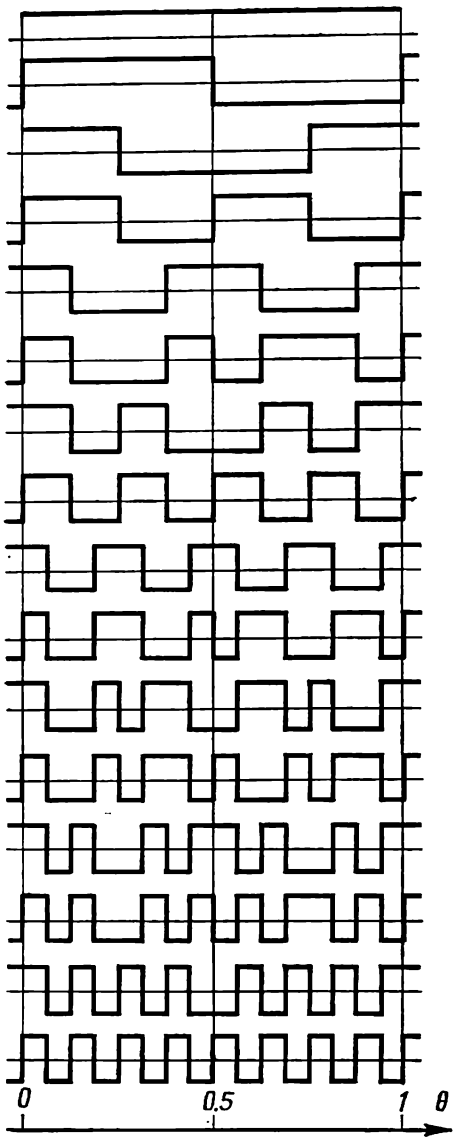
with coefficients

$$A(i) = \int_0^1 f(\theta)\ \text{wal}\ (i,\ \theta)\, d\theta,\ \theta = t/T \qquad (14.39)$$

Outside the interval (0, 1), series (14.38) describes a periodic function $f(\theta + k)$, where k is any integer.

Figure 14.15 shows the first sixteen Walsh functions.

Some specific features of the expansion of continuous functions in sets of the Walsh functions are illustrated by examples in Sec. 14.5.



wal(w,θ) w	pal(p,θ) p	had(h,θ) h
0	0	0
1	1	8
2	3	12
3	2	4
4	6	6
5	7	14
6	5	10
7	4	2
8	12	3
9	13	11
10	15	15
11	14	7
12	10	5
13	11	13
14	9	9
15	8	1

Fig. 14.15. Numbering of the Walsh functions with different methods of ordering. The basis $N = 16$

As mentioned before, the Walsh functions are convenient for modern microelectronics and can easily be formed by means of switching circuits. One possible version of a generator for generating the first eight functions is shown in Fig. 14.16.

In this generator, the algorithm for forming the Walsh functions is based on expression (14.30), i.e., on the multiplication of the like

powers of three Rademacher functions: $r_1(\theta)$, $r_2(\theta)$, and $r_3(\theta)$. The function $r_3(\theta)$ is obtained directly from a square-wave generator. The second function $r_2(\theta)$ is produced from $r_3(\theta)$ by doubling the period of this wave. This is effected by a flip-flop started by the

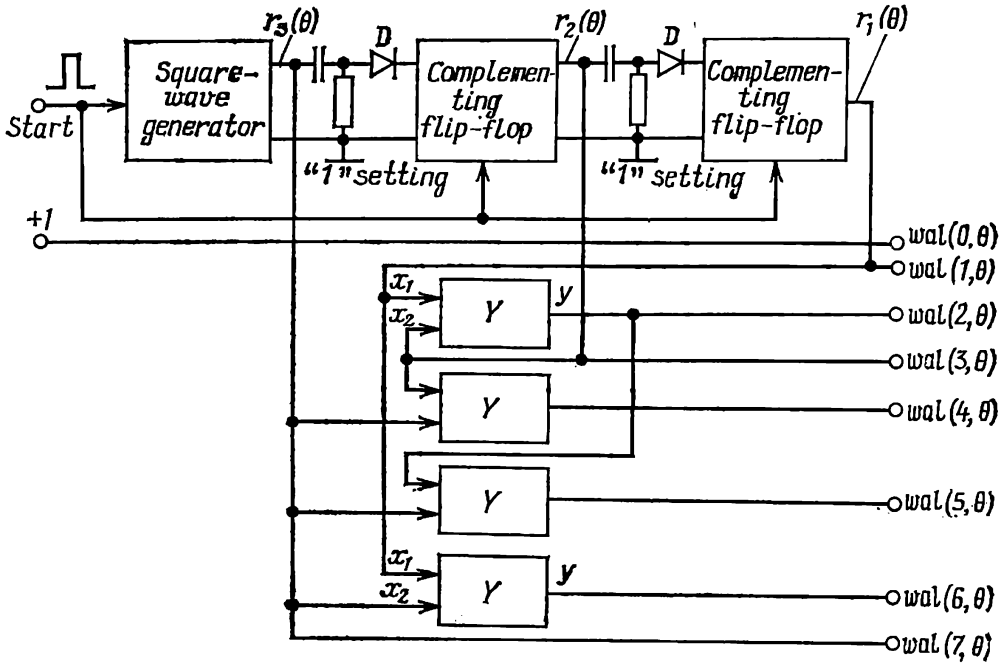


Fig. 14.16. Generator of the first eight Walsh functions

positive pulses (taken from the output of a differentiating R-C circuit and a diode) arising at the beginning of each period of the input square wave. The function $r_1(\theta)$ is obtained by a similar way from $r_2(\theta)$. The functions $r_1(\theta)$, $r_2(\theta)$, and $r_3(\theta)$ thus obtained are applied either directly to the output of the generator or to the multipliers Y. Each multiplier is a coincidence gate having the following truth table:

Inputs		Output
x_1	x_2	y
0	0	1
0	1	0
1	0	0
1	1	1

The device is started by a pulse which sets all the flip-flops to the "1" state simultaneously.

14.5. EXAMPLES OF APPLICATION OF THE WALSH FUNCTIONS

1. Determining the spectrum of a sinusoid $s(t) = \sin \frac{2\pi}{T} t$ (Fig. 14.17*a*) in the basic set of Walsh functions. In this case, the expansion interval T_0 is preferably taken equal to T. Transforming to dimensionless time $\theta = t/T$, let us write the oscillation $s(t)$ in the form $s_1(\theta) = \sin 2\pi\theta$. We shall restrict ourselves to 16 functions and, at first, we shall use the ordering by Walsh. Since the given function $s_1(\theta)$ is odd with respect to the point $\theta = 1/2$, all the coefficients $A(i)$ multiplying the even Walsh functions in series (14.38), i.e., those multiplying the functions cal (j, θ), are zero.

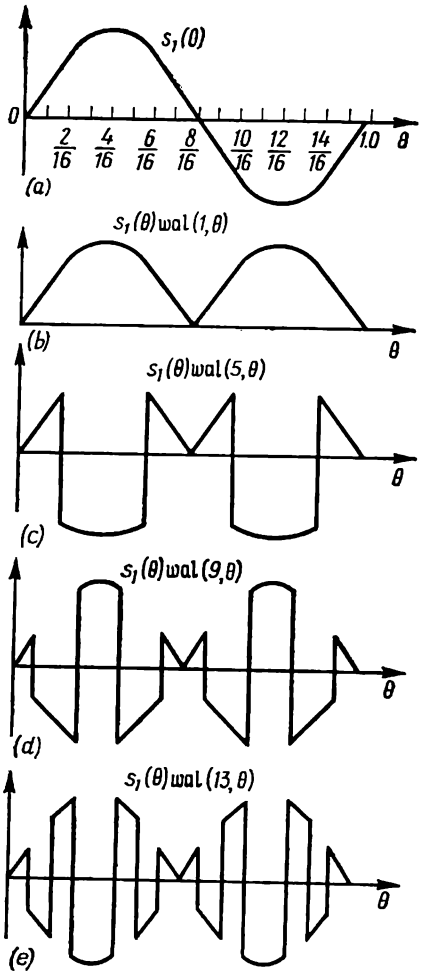


Fig. 14.17. Sampling of a section of a sine curve by the Walsh functions

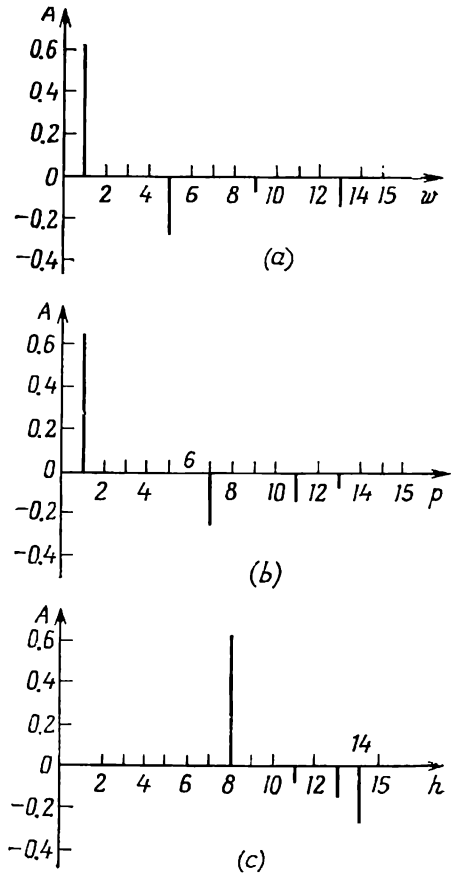


Fig. 14.18. Spectra of a sine curve in the basic set of Walsh functions ordered by (*a*) Walsh, (*b*) Paley, and (*c*) Hadamard. The basis $N = 16$

Those of the remaining eight functions wal (i, θ), which coincide with the Rademacher functions and whose period within the interval $[0, 1[$ is a multiple of the period of the function $s_1(\theta)$, also have zero coefficients $A(i)$. Such functions include wal $(3, \theta)$, wal $(7, \theta)$, and wal $(15, \theta)$. Finally, the function wal $(11, \theta)$ which is odd not only relative to the point $\theta = 1/2$ but also relative to the points $\theta = 1/4$ and $\theta = 3/4$ [within the intervals $[0, 1/2[$ and $(1/2, 1)]$ has zero coefficient $A(11)$ as well, because $s_1(\theta)$ is even within these intervals.

Thus, only four of the sixteen coefficients are nonzero: $A(1)$, $A(5)$, $A(9)$ and $A(13)$. Let us find these coefficients using formula (14.39). The integrands, which are the products of the signal $s_1(\theta)$ (Fig. 14.17*a*) and the corresponding function wal (i, θ), are shown in Fig. 14.17*b* through *e*. Piecewise integration of these functions yields

$$A(1) = 2\int_0^{1/2} \sin 2\pi\theta\, d\theta = 2/\pi = 0.636$$

$$A(5) = 4\int_0^{2/16} \sin 2\pi\theta\, d\theta - 2\int_{2/16}^{6/16} \sin 2\pi\theta\, d\theta = \frac{2}{\pi}\left(1 - 2\cos\frac{\pi}{4}\right) = -0.265$$

$$A(9) = 4\int_0^{1/16} \sin 2\pi\theta\, d\theta - 4\int_{1/16}^{3/16} \sin 2\pi\theta\, d\theta + 2\int_{3/16}^{5/16} \sin 2\pi\theta\, d\theta = -0.052$$

$$A(13) = 4\int_0^{1/16} \sin 2\pi\theta\, d\theta - 4\int_{1/16}^{2/16} \sin 2\pi\theta\, d\theta + 4\int_{2/16}^{3/16} \sin 2\pi\theta\, d\theta - 2\int_{3/16}^{5/16} \sin 2\pi\theta\, d\theta = -0.128$$

The spectrum of the signal $s_1(\theta)$ in the basic set of Walsh functions (Walsh-ordered) is shown in Fig. 14.18*a*. When ordered by Paley and by Hadamard, the spectrum of the same signal assumes the form shown in Fig. 14.18*b* and *c*. These spectra are obtained from the spectrum in Fig. 14.18*a* by transposing the coefficients in accordance with the table (Fig. 14.15) illustrating the interrelation between the methods of ordering of the Walsh functions (for $N = 16$).

To reduce distortion when representing an oscillation by a limited number of Walsh functions, preference should be given to the ordering that provides for a monotonic decrease of the spectrum. In other words, the best is the ordering with which each next component of the spectrum has a magnitude not exceeding the magnitude of the previous one, i.e., $|A(i+1)| \leqslant |A(i)|$. In this sense, the best order-

ing for a section of a sine curve, as it follows from Fig. 14.18, is the ordering by Paley and the worst one is that by Hadamard.

The restoration of the original signal (Fig. 14.17*a*) by means of four Walsh functions is shown in Fig. 14.19. Of course, this plotting

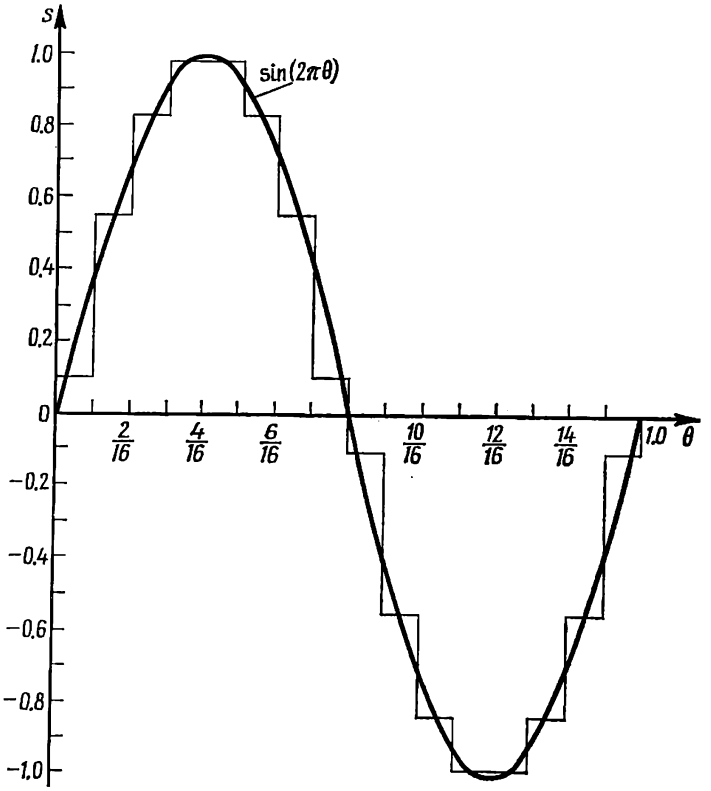


Fig. 14.19. Approximation of a sine curve by the Walsh functions

is independent of the method of ordering the functions. Satisfactory approximation of a sine wave in the basic set of Walsh functions requires a substantial increase in the number of the spectral components.

As mentioned in Sec. 14.4, outside the interval [0, 1[, series (14.38) describes a periodic continuation $s_1(\theta)$ which in the given example is a harmonic function.

2. The spectrum of a harmonic oscillation $s(t) = \cos(\omega t + \theta_0)$ (Fig. 14.20) in the basic set of Walsh functions. As in the preceding example, we shall consider one cycle of a harmonic oscillation with a period $T = 2\pi/\omega$. Transforming to dimensionless time $\theta = t/T$, we write the oscillation $s(t)$ in the form

$$s_1(\theta) = \cos(2\pi\theta + \theta_0) = \cos\theta_0 \cos 2\pi\theta - \sin\theta_0 \sin 2\pi\theta$$
$$= A \cos 2\pi\theta - B \sin 2\pi\theta$$

The Walsh spectrum of the function $\sin 2\pi\theta$ is defined in Example 1. The spectrum of the function $\cos 2\pi\theta$ within the interval $[0, 1[$ is determined in a similar way. It is only necessary to replace

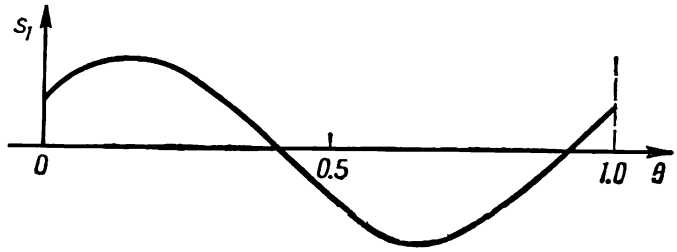


Fig. 14.20. One period of a harmonic oscillation within the interval $0 \leqslant \theta < 1$

the functions sal (j, θ) by the functions cal (j, θ). It can be easily verified that in the case of the ordering by Walsh the new coefficients $A(i)$ in series (14.38) will be as follows: $A(2)$, $A(6)$, $A(10)$ and $A(14)$ instead of $A(1)$, $A(5)$, $A(9)$ and $A(13)$, the magnitudes of the coefficients remaining the same.

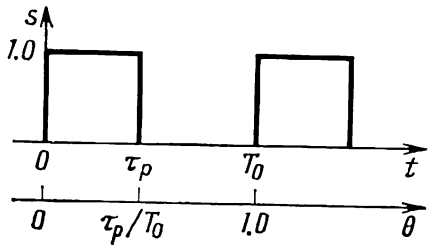


Fig. 14.21. One cycle of a periodic pulse train with $\tau_p/T_0 = 1/2$

Thus, series (14.38) for the oscillation under consideration can be written in the form

$$\begin{aligned} s_1(\theta) &= \cos\theta_0 \cos 2\pi\theta \\ &\quad - \sin\theta_0 \sin 2\pi\theta \\ &= \cos\theta_0 \,[A(2)\,\mathrm{wal}\,(2,\theta) \\ &\quad + A(6)\,\mathrm{wal}\,(6,\theta) \\ &\quad + A(10)\,\mathrm{wal}\,(10,\theta) + A(14)\,\mathrm{wal}\,(14,\theta)] \\ &\quad - \sin\theta_0 \,[A(1)\,\mathrm{wal}\,(1,\theta) + A(5)\,\mathrm{wal}\,(5,\theta) \\ &\quad + A(9)\,\mathrm{wal}\,(9,\theta) + A(13)\,\mathrm{wal}\,(13,\theta)] \end{aligned}$$

Thus, when a harmonic oscillation is shifted in phase, the Walsh spectrum comprises even and odd functions cal (j, θ) and sal (j, θ).

3. The spectrum of a periodic train of square pulses (Fig. 14.21) in the basic set of Walsh functions. Let us define the oscillation $s(t)$ within the interval $(0, T_0)$ by the expression $s(t) = 1$, $0 \leqslant t < \tau_p$, and respectively,

$$s_1(\theta) = \begin{cases} 1 & \text{for } 0 \leqslant \theta < \tau_p/T_0 \\ 0 & \text{for } \theta \geqslant \tau_p/T_0 \end{cases} \qquad (14.40)$$

The structure of the Walsh spectrum of the given oscillation greatly depends on the relation between τ_p and T_0. The time base T_0 is an additional and arbitrarily selected parameter of the Walsh func-

tions. In fact, for $\tau_p/T_0 = 1$, the spectrum contains only one function wal (0, θ) with the coefficient A (0) = 1. For $\tau_p/T_0 = 1/2$, oscillation (14.40) is completely defined by two functions: wal (0, θ) and wal (1, θ) with the coefficients A (0) = A (1) = 1/2.

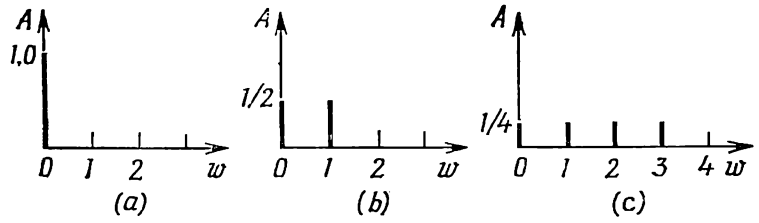


Fig. 14.22. Spectra of a train of square pulses in the basic set of Walsh functions with

(a) $\tau_p/T_0 = 1$, (b) $\tau_p/T_0 = 1/2$ and (c) $\tau_p/T_0 = 1.4$. $N = 16$

Further, for $\tau_p/T_0 = 1/4$, the use of formula (14.39)

$$A(w) = \int_0^1 s(\theta)\ \text{wal}\ (w, \theta)\, d\theta = \int_0^{1/4} \text{wal}\ (w, \theta)\, d\theta$$

gives the following coefficients: A (0) = A (1) = A (2) = A (3) = = 1/4.

The spectra thus found are shown in Fig. 14.22. This result is easily generalized for a train of square pulses with the ratio $\tau_p/T_0 = 1/2^k$, where k is a positive integer. It is clear that the Walsh spectrum of such a wave consists of 2^k components with identical amplitudes equal to $1/2^k$. It is very important that this spectrum includes a finite number of components; the expansion of the same oscillation (14.40) in harmonic function is infinite.

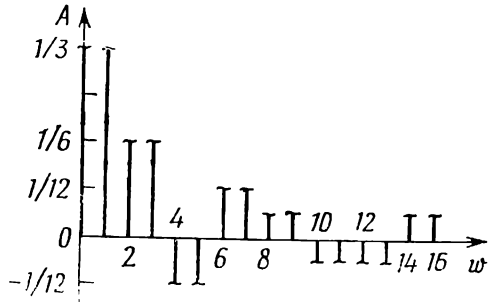


Fig. 14.23. Spectrum of a train of square pulses in the basic set of Walsh functions with $\tau_p/T_0 = 1/3$ and $N = 16$

Now, let us consider the case where $\tau_p/T_0 \neq 1/2^k$, e.g., $\tau_p/T_0 = 1/3$. Restricting ourselves to the first 16 Walsh functions (Walsh-ordered) and omitting intermediate manipulations, we obtain

$$A(0) = A(1) = 1/3, \quad -A(4) = -A(5) = A(6) = A(7) = 1/12$$
$$A(8) = A(9) = -A(10) = -A(11) = -A(13)$$
$$= A(14) = A(15) = 1/24$$

The spectrum thus found is shown in Fig. 14.23. When transformed to the ordering by Paley, the spectrum structure is preserved (in magnitudes).

Thus, for $\tau_p/T_0 \neq 1/2^k$, the Walsh spectrum of a periodic train of square pulses includes an infinitely large number of compo-

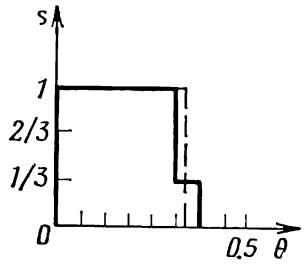


Fig. 14.24. Approximation of a square pulse by 16 Walsh functions with $\tau_p/T_0 = 1/3$ and $N = 16$

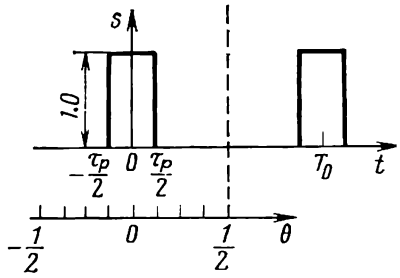


Fig. 14.25. One cycle of a periodic train of pulses within the interval $0.5 \leqslant \theta < 1$

nents. The sum of the first 16 functions gives a pulse shown in Fig. 14.24.

4. The effect of a shift of a periodic train of pulses on the Walsh spectrum.

Let us consider this question, using as an example a pulse train with $\tau_p/T_0 = 1/4$ (Fig. 14.25), shifted by an amount $\tau_p/2$ with respect to a similar train, see Example 3.

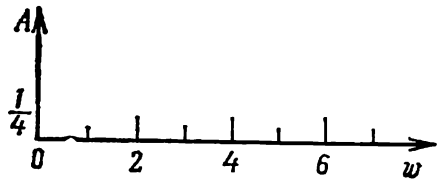


Fig. 14.26. Effect of a shift of a pulse train on the Walsh spectrum (compare with Fig. 14.22*c*)

Using the Walsh functions (Walsh-ordered) determined within the interval $-1/2 \leqslant \theta < 1/2$ (see Fig. 14.14), let us write the expression for the Walsh coefficients

$$A(w) = \int_{-1/8}^{1/8} \text{wal}(w, \theta)\, d\theta$$

from which we obtain the following nonzero coefficients:

$$A(0) = A(2) = A(4) = A(6) = 1/4$$

The spectrum thus obtained (Fig. 14.26) is twice as wide as that shown in Fig. 14.22*c*. Thus, the shift of the pulse train for the time $\tau_p/2$ has resulted in a change of the spectrum. The dependence of the spectrum structure on the shift of the oscillation $s(t)$ along the time axis is a specific feature of the analysis in the basic set of Walsh functions. This feature is associated with the nonperiodicity of the Walsh functions within the unit interval of their definition. It should be recalled that in the case of expansion in harmonic functions, the time shift of the signal influences only the phase-frequency characteristic of the spectrum (see Sec. 2.7).

14.6. MUTUAL SPECTRUM OF TWO DIFFERENT COMPLETE ORTHONORMAL SETS OF FUNCTIONS

Let us expand a given oscillation $s(t)$ in terms of two complete orthonormal sets of functions $\varphi_k(t)$ and $\xi_n(t)$ with identical definition intervals (a, b):

$$s(t) = \sum_{k=0}^{\infty} c_k \varphi_k(t) = \sum_{n=0}^{\infty} b_n \xi_n(t) \text{ for } a \leqslant t < b \qquad (14.41)$$

The set of coefficients c_k forms the spectrum of the oscillation $s(t)$ in the orthonormal basis $\varphi(t)$, while that of the coefficients b_n forms the spectrum of the same oscillation in the orthonormal basis $\xi(t)$.

Then, let us use expression (2.9) for the coefficients c_k of the generalized Fourier series and substitute into this expression the given function $s(t)$ [instead of $f(x)$] expressed through the right-hand side of (14.41):

$$c_k = \int_a^b s(t)\,\varphi_k(t)\,dt = \int_a^b \left[\sum_{n=0}^{\infty} b_n \xi_n(t)\right] \varphi_k(t)\,dt$$

$$= \sum_{n=0}^{\infty} b_n \int_a^b \xi_n(t)\,\varphi_k(t)\,dt \qquad (14.42)$$

Owing to the orthonormal nature of the functions $\varphi_k(t)$, the coefficient multiplying the integral in (2.9) is taken equal to unity. The integral under the sum sign is the kth coefficient of expansion of the function $\xi_n(t)$ in the basic set of functions $\varphi_k(t)$. Denoting this coefficient by the symbol

$$d_{\xi_n \varphi_k} = \int_a^b \xi_n(t)\,\varphi_k(t)\,dt \qquad (14.43)$$

we obtain

$$c_k = \sum_{n=0}^{\infty} b_n d_{\xi_n \varphi_k} \tag{14.44}$$

The coefficients $d_{\xi_n \varphi_k}$ are independent of the signal $s(t)$. Thus, expression (14.44) establishes the relation between the spectra of the oscillation $s(t)$ expanded in two different complete orthonormal sets of functions $\varphi_k(t)$ and $\xi_n(t)$.

Let us consider expression (14.43) in more detail. It can be equally treated either as the kth coefficient of the expansion of the function

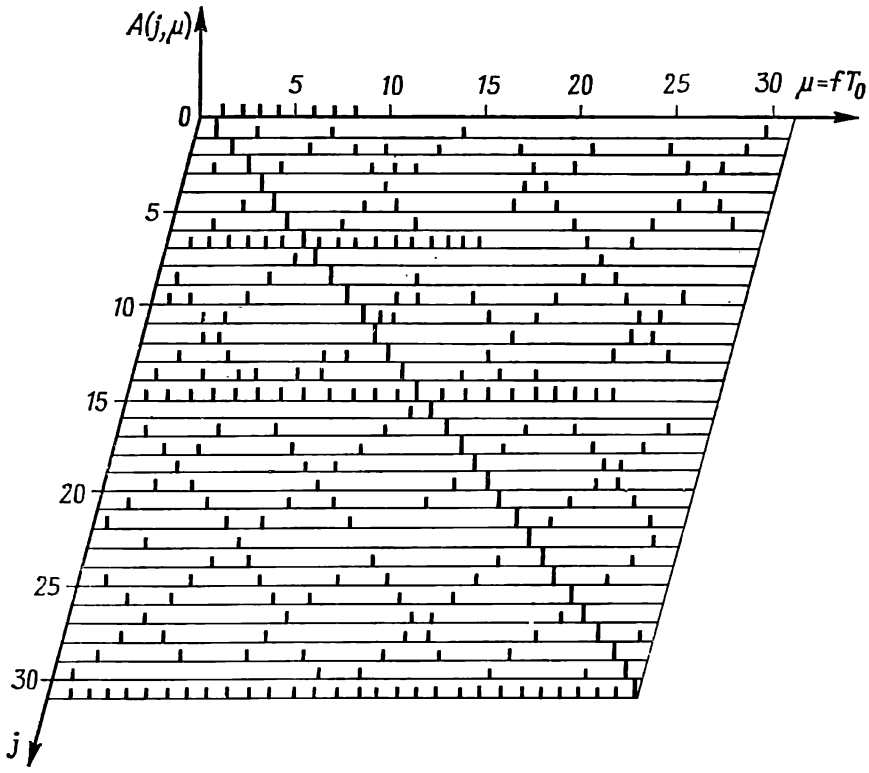


Fig. 14.27. Mutual spectrum of Walsh functions and trigonometric functions (even components)

$\xi_n(t)$ in the complete orthonormal set of functions $\varphi(t)$ or as the nth coefficient of the expansion of the function $\varphi_k(t)$ in the complete orthonormal set of functions $\xi(t)$. Therefore, the set of the coefficients $d_{\xi_n \varphi_k}$ may be regarded as the *mutual spectrum* of the basic sets of functions $\varphi_k(t)$ and $\xi_n(t)$.

As an example, let us consider the mutual spectrum of the Walsh functions and trigonometric functions. Using expression (14.43), let us write similar expressions separately for the even and the odd

functions defined within the interval [—1/2, 1/2[:

$$B(j,\ \mu) = \int_{-1/2}^{1/2} \mathrm{sal}\,(j,\ \theta) \sin(2\pi\mu\theta)\, d\theta \tag{14.45}$$

$$A(j,\ \mu) = \int_{-1/2}^{1/2} \mathrm{cal}\,(j,\ \theta) \cos(2\pi\mu\theta)\, d\theta \tag{14.46}$$

The arguments ωt of the trigonometric functions are reduced to the form

$$\omega t = \frac{2\pi}{T} t = 2\pi f T_0 \frac{t}{T_0} = 2\pi\mu\theta, \quad \mu = fT_0 \tag{14.47}$$

Figure 14.27 shows the relief of the modulus of the mutual spectrum of the even Walsh functions ($N = 64$) and sine functions. It is clear that the relief of the mutual spectrum $B\ (j,\ \mu)$ is characterized by a well-pronounced symmetry. The maxima of $|B\ (j,\ \mu)|$ are near the plane $j = \mu$. The additional planes of symmetry connecting the side maxima are normal to the main plane and intersect it at intervals 2^k at the points $j = \mu = 2^k$, $k = 0, 1, 2, \ldots$

14.7. DISCRETE WALSH FUNCTIONS

The discrete Walsh functions are of the greatest interest for the digital methods of spectral analysis. These functions are samples of the continuous Walsh functions. To transform to the discrete version of the functions, let us represent the argument $0 \leqslant \theta < 1$ in the binary form

$$\theta = \sum_{k=1}^{\infty} \theta_k 2^{-k} = (0,\ \theta_1,\ \theta_2,\ \ldots)_2 \tag{14.48}$$

where $\theta_k = 0;\ 1,\ k = 1, 2, 3, \ldots$

As θ changes from zero to unity, the kth digit θ_k alternately takes on values 0 or 1, the interval between two adjacent sign alternations being equal to $1/2^k$. Thus, for $k = 1$, the digit θ_1 is zero within the interval $0 \leqslant \theta < 1/2$ and is unity within the interval $1/2 \leqslant \theta < 1$. For $k = 2$, $\theta = 0$ within the intervals [0, 1/4], (1/2, 3/4) and $\theta_2 = 1$ within the intervals (1/4, 1/2), (3/4, 1), etc.

Therefore, expression (14.26) for the kth Rademacher function can be written in the form

$$r_k(\theta) = e^{i\pi\theta_k} = (-1)^{\theta_k} \tag{14.49}$$

Substituting this expression into (14.30), we obtain the following expression for the Walsh functions (Walsh-ordered):

$$\mathrm{wal}\ (w,\theta) = \prod_{k=1}^{n} [e^{i\pi\theta_k}]^{w_{n-k+1} \oplus w_{n-k}} = (-1)^{\sum_{k=1}^{n} (w_{n-k+1} \oplus w_{n-k})\theta_k} \tag{14.50}$$

The result will not change if the argument $0 \leqslant \theta < 1$ is replaced by an integral variable $x = 0, 1, \ldots, N-1$, which is the serial sample number and is written in the binary form

$$x = \sum_{k=1}^{n} x_k 2^{n-k} = (x_1, x_2 \ldots x_n)_2 \tag{14.51}$$

where $x_k = 0; 1$, $k = 1, 2, 3, \ldots, N = 2^n$.

Then expression (14.50) assumes the following form:

$$\operatorname{wal}(w, x) = (-1)^{\sum_{k=1}^{n} (w_{n-k+1} \oplus w_{n-k}) x_k} \tag{14.52}$$

The first two and the last $(N-1)$th discrete Walsh functions determined by expression (14.52) are shown in Fig. 14.28 (for $N = 16$). Each sample is located in the middle of the associated element of the continuous function. The duration of the element is equal to $1/N$ of the interval $(0, 1)$.

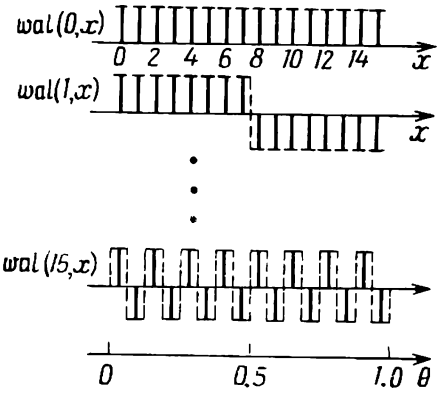


Fig. 14.28. First two and the last discrete Walsh functions with $N = 2^3 = 16$ (ordered by Walsh)

Taking into account (14.32) and (14.34), the discrete Walsh functions ordered by Paley and Hadamard are written in the form

$$\left.\begin{aligned} \operatorname{pal}(p, x) &= (-1)^{\sum_{k=1}^{n} p_{n-k+1} x_k} \\ \operatorname{had}(h, x) &= (-1)^{\sum_{k=1}^{n} h_k x_k} \end{aligned}\right\} \tag{14.53}$$

The properties of the continuous Walsh functions, given in Sec. 14.4, are written for the discrete functions in the following way.

Orthogonality

$$\sum_{x=0}^{N-1} \operatorname{wal}(i, x) \operatorname{wal}(l, x) = \begin{cases} N & \text{if } i = l \\ 0 & \text{if } i \neq l \end{cases} \tag{14.54}$$

The discrete Walsh functions are not normalized; the norm is equal to N regardless of the serial number of the function.

Multiplicativity

$$\operatorname{wal}(i, x) \operatorname{wal}(l, x) = \operatorname{wal}(i \oplus l, x) \tag{14.55}$$

Let a signal $s(t)$ (a real function) be represented by a set of its equidistant samples $s(k)$, $k = 0, 1, 2, \ldots, N-1$. Then the trans-

forms

$$S(i) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) \operatorname{wal}(i, k) \tag{14.56}$$

$$s(k) = \sum_{n=0}^{N-1} S(i) \operatorname{wal}(i, k) \tag{14.57}$$

form a pair of *discrete Walsh transforms*. Expressions (14.56) and (14.57) are similar to the pair of discrete Fourier transforms in basic sets of harmonic functions [see (13.13′) and (13.14)].

Like the discrete Fourier transforms [see Sec. 13.3], the discrete Walsh transforms are periodic:

$$S(n) = S(n + mN), \; s(k) = s(k + mN) \tag{14.58}$$

where m is an integer.

However, the discrete Walsh transforms have some specific features as to the delay theorem. It should be recalled that in the case of spectral analysis in basic sets of harmonic functions, the multiplication of the discrete Fourier transform $\mathbf{S}_T(n\omega_1)$ by the basic function $e^{i\frac{2\pi}{N}nm}$ is equivalent to a time shift of the sequence $s(k)$, $k = 0, 1, 2, \ldots, N-1$, for m intervals (see Sec. 13.3).

In the case of the discrete Walsh transforms, the multiplication of both parts of equation (14.56) by the basic function $\operatorname{wal}(n, m)$ yields

$$\operatorname{wal}(n, m) S(n) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) \operatorname{wal}(n, k) \operatorname{wal}(n, m) \tag{14.59}$$

which can be written in the form

$$\operatorname{wal}(n, m) S(n) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) \operatorname{wal}(n, k \oplus m)$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} s(k \oplus m) \operatorname{wal}(n, k) \tag{14.60}$$

because of the multiplicative properties of the Walsh functions [see (14.36) and (14.37)].

Transformation from $s(k)$ to $s(k \oplus m)$ implies a dyadic shift of the sequence of samples $s(k)$, $k = 0, 1, 2, \ldots, N-1$, for m intervals.

In a similar way, we may show that the dyadic shift of the spectral components $S(n)$ for m intervals corresponds to the multiplication of the samples $s(k)$ by $\operatorname{wal}(k, m)$.

Let us clarify the meaning of the term "dyadic shift". One has to deal with the concept of shift of a function, for example, when calculating correlation functions, when studying the delay theorem, when determining the convolution of two functions. In the common sense, shift is regarded as the parallel transfer of the values of an oscillation along the time axis. Such a shift may be called arithmetic, because it is usually expressed by ordinary arithmetic addition or subtraction. For example, in the case of arithmetic shift for $m = 3$ intervals, the kth sample $x(5)$ will be shifted and will become $x(5 + 3) = x(8)$. With sufficiently large m, the sample $x(k)$ will come beyond the limits of the original set of samples. In the case of dyadic shift, the same sample $x(5)$ shifted for $m = 3$ will occupy the position $x(5 \oplus 3) = x(6)$, because

$$\begin{array}{r} 5 = (101)_2 \\ \oplus \\ 3 = (011)_2 \\ \hline (110)_2 = 6 \end{array}$$

The dyadic shift features the so-called *group property*: the shift of samples $x(k)$ (where $k = 0, 1, 2, \ldots, N - 1$) by an amount $m \leqslant N - 1$ corresponds to such a transposition of these samples *inside* their original set that the result of the digit-by-digit modulo 2 addition, i.e., $k \oplus m$, never exceeds the number $N - 1$ irrespective of $m = 0, 1, 2, \ldots, N - 1$, it being implied that $N = 2^n$, where n is a positive integer. This statement can be easily verified by taking all possible dyadic shifts of all samples $x(k)$ with a given N. For example, for $N = 8$, we have the following square matrix of the values $q = k \oplus m$:

$$q = \begin{array}{c|cccccccc} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 0 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 1 & 0 & 3 & 2 & 5 & 4 & 7 & 6 \\ 2 & 2 & 3 & 0 & 1 & 6 & 7 & 4 & 5 \\ 3 & 3 & 2 & 1 & 0 & 7 & 6 & 5 & 4 \\ 4 & 4 & 5 & 6 & 7 & 0 & 1 & 2 & 3 \\ 5 & 5 & 4 & 7 & 6 & 1 & 0 & 3 & 2 \\ 6 & 6 & 7 & 4 & 5 & 2 & 3 & 0 & 1 \\ 7 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \end{array} \quad \begin{array}{l} k \\ \\ \\ \\ \\ \\ \\ \\ \\ m \end{array}$$

From this matrix it is clear that the dyadic shift does not bring the samples being shifted outside the original set of N samples but merely causes their transposition within this set.

The dyadic shift imparts specific features both to the spectral analysis in the basic set of Walsh functions and to the representation of signals in the time domain. In particular, a dyadic convolution of two time sequences $x(k)$ and $y(k)$ is written in the form

$$h(k) = \frac{1}{N} \sum_{m=0}^{N-1} x(m)\, y(k \oplus m) = x(k) \ast y(k) \qquad (14.61)$$

The main advantage of the discrete Walsh transforms over the discrete Fourier transforms is that the signal samples are multiplied by the Walsh functions, which take on the values ± 1 [see (14.56) and (14.57)]. In essence, the operation of multiplication is excluded, and expressions (14.56) and (14.57) are reduced to addition of the samples with the respective signs, while in the case of the discrete Fourier transforms, multiplication by complex numbers of the form $e^{\pm i \frac{2\pi}{N} nk}$ is required, the real and imaginary parts of these numbers requiring representation by an adequately large number of digits (to reduce the rounding-off error).

The efficiency of the digital methods of spectral analysis can be increased considerably by using special transform algorithms known as the *fast Fourier transform*, the *fast Walsh transform*, etc. These algorithms are based on the elimination of the redundancy in the description of the matrix of a discrete orthogonal transform.

Chapter 15

ELEMENTS OF SYNTHESIS OF LINEAR RADIO CIRCUITS

15.1. GENERAL

The classical theory of synthesis of passive linear electric circuits with lumped parameters envisages two stages:

—selection of an appropriate rational function which could be a characteristic of a physically realizable circuit and, at the same time, be sufficiently close to the specified characteristic;

—finding the structure and elements of the circuit realizing the selected function.

The first stage is known as the approximation of the given characteristic and the second, as the realization of the circuit.

The approximation, based on the use of various orthogonal functions, causes no principal difficulties. The problem of finding the optimum structure of the circuit by the given (physically realizable) characteristic of this circuit is much more complicated. This problem has no single-valued solution. The same characteristic of a circuit can be realized by many methods differing in schematic solutions, the number of circuit elements and their complexity, the sensitivity of the circuit characteristics to the instability of the parameters, etc.

Known in the art is synthesis of circuits in the frequency domain and in the time domain. In the first case, the transfer function $\mathbf{K}(i\omega)$ is specified and in the second, the impulse response $g(t)$. Since these two functions are interrelated by a pair of Fourier transforms, the synthesis of circuits in the time domain can be reduced to the synthesis in the frequency domain and vice versa. However, the synthesis by a given impulses response has its specific features playing an important role when shaping pulses with definite requirements for their parameters (steepness of the leading edge, spike, shape of the top, etc.).

In this chapter we shall consider the synthesis of two-port networks in the frequency domain. It should be noted that a vast literature is presently available on the synthesis of linear electric circuits, and the study of the general synthesis theory is not covered in this course. Here we discuss only some particular questions of the synthesis of two-port networks, that reflect some characteristic features of modern radioelectronic circuits, such as the use of active two-port networks, trends towards the elimination of inductances from selective circuits (in microcircuits) and intensive development of discrete circuit engineering.

It is known that the transfer function $\mathbf{K}(i\omega)$ of a two-port network is unambiguously determined by its zeros and poles on the p-plane. Therefore, the expression "synthesis by the given transfer function" is equivalent to the expression "synthesis by the given zeros and poles of the transfer function". The existing theory of synthesis of two-port networks concerns circuits whose transfer function has a *finite* number of zeros and poles, i.e., circuits consisting of a finite number of links with lumped parameters. From this stems the conclusion that the classical methods of synthesis of circuits are inapplicable to filters matched to a given signal. In fact, the factor $e^{-i\omega t}$ [see (12.16)] entering into the transfer function of such a filter cannot be realized by a finite number of links with lumped parameters. The material presented in this chapter is oriented towards two-port networks with few links. Such two-port networks are typical of low-pass filters, high-pass filters, rejection filters, etc., which are widely used in radioelectronic devices.

Sections 15.2 through 15.4 outline the fundamentals of the transfer functions of two-port networks, that are required for understanding the methods of their synthesis.

15.2. SOME PROPERTIES OF THE TRANSFER FUNCTION OF A TWO-PORT NETWORK

It is known from the theory of linear electric circuits that the transfer function of a physical two-port network with lumped parameters can be represented as a rational fraction — the ratio of two polynomials with integral powers of p and with real coefficients

$$\mathbf{K}(p)=\frac{a_0p^n+a_1p^{n-1}+\ldots+a_n}{b_0p^m+b_1p^{m-1}+\ldots+b_m}=\frac{P(p)}{Q(p)} \qquad (15.1)$$

with some limitations as to the position of the roots of the polynomials $P(p)$ and $Q(p)$ on the p-plane and also the relation between the powers n and m.

A general requirement for the transfer function $\mathbf{K}(p)$ of any two-port network, which stems from the condition of stability, is the absence of poles in the right-hand half-plane p (see Sec. 5.10). For this purpose, it is necessary that the roots of the polynomial $Q(p)$ in the denominator of expression (15.1) be located only in the left-hand half-plane p. Such polynomials are called the *Hurwitz polynomials.*

Further, in order to ensure that rational fraction (15.1) is the transfer function of a physical two-port network, the power n of the numerator must not exceed the power m of the denominator.

Additional information on the number of zeros and poles of the function $\mathbf{K}(p)$ and on their location on the p-plane can be obtained

from expression (15.1) written in the form

$$\mathbf{K}(p) = \frac{a_0 (p - p_{01})(p - p_{02}) \dots (p - p_{0n})}{b_0 (p - p_{p1})(p - p_{p2}) \dots (p - p_{pm})} = \frac{a_0 \prod_{k=1}^{n} (p - p_{0k})}{b_0 \prod_{k=1}^{m} (p - p_{pk})} \qquad (15.2)$$

where p_{0k} are the zeros of the function $\mathbf{K}(p)$ [the roots of the equation $P(p) = 0$] and p_{pk} are the poles [the roots of the equation $Q(p) = 0$].

If $n = m$, the function $\mathbf{K}(p)$ has n zeros and n poles for p equal to the roots of the equations $P(p) = 0$ and $Q(p) = 0$, respectively. If $n < m$, the function $\mathbf{K}(p)$ has n zeros at the p-plane points corresponding to the n roots of the equation $P(p) = 0$ and, in addition, $m - n$ zeros at the point $p = \infty$. In fact, as $p \to \infty$, expression (15.2) turns into

$$\lim_{p \to \infty} \mathbf{K}(p) = \lim_{p \to \infty} \frac{a_0}{b_0} p^{(n-m)} = \lim_{p \to \infty} \frac{a_0}{b_0} \frac{1}{p^{m-n}} = 0$$

i.e., the point $p = \infty$ is the zero of multiplicity $(m - n)$.

The total number of zeros (finite and infinite) is equal to m. Thus, taking into account the zeros at infinity, the rational function $\mathbf{K}(p)$ has the same number of zeros and poles.

The zeros and poles may be either real or complex. In the latter case, the zeros or poles form complex-conjugate pairs. Expression (15.2), if there are no real zeros and poles, may be written in the form

$$\mathbf{K}(p) = \frac{a_0}{b_0} \frac{(p - p_{01})(p - p_{01}^*)(p - p_{03})(p - p_{03}^*) \dots (p - p_{0(n-1)})(p - p_{0(n-1)}^*)}{(p - p_{p1})(p - p_{p1}^*)(p - p_{p3})(p - p_{p3}^*) \dots (p - p_{p(m-1)})(p - p_{p(m-1)}^*)} \qquad (15.3)$$

In contrast to the poles, the zeros of the transfer function can be located both in the left-hand and the right-hand half-planes p. The two-port networks with the zeros of their transfer functions in the right-hand half-plane have essential specific features which are discussed in the next section.

15.3. RELATION BETWEEN THE AMPLITUDE-FREQUENCY AND PHASE-FREQUENCY CHARACTERISTICS OF A TWO-PORT NETWORK

The analysis of transmission of an oscillation through a linear circuit is based on the use of the amplitude-frequency and phase-frequency characteristics of this circuit. A question arises whether it is possible to control one characteristic without changing the other, or there is a one-to-one relation between them? This question

is of great practical and scientific interest for radioelectronics, particularly as concerns the synthesis of circuits by the given amplitude frequency and phase-frequency characteristics.

The problem is reduced to finding the relation between the modulus and argument of the complex function $\mathbf{K}(p)$ having the following properties: (a) the number of poles is finite; (b) there is no poles in the right-hand half-plane p. This question leads to one of the most complex problems of the theory of functions of a complex variable. A simpler problem is to express the real part of the transfer function through the imaginary part or vice versa. Therefore, it is expedient to study not the function $\mathbf{K}(p)$, but the function $\Theta(p)$ related to $\mathbf{K}(p)$ by the relation

$$\Theta(p) = \ln \mathbf{K}(p) \tag{15.4}$$

On the frequency axis, this new function assumes the form

$$\Theta(i\omega) = \ln \mathbf{K}(i\omega) = \ln [K(\omega) e^{i\varphi(\omega)}]$$

$$= \ln K(\omega) + i\varphi(\omega) = A(\omega) + i\varphi(\omega) \tag{15.5}$$

The real part of this function

$$A(\omega) = \ln K(\omega) \tag{15.6}$$

is called the logarithmic attenuation of the two-port network.

Taking into account that

$$K(\omega) = e^{A(\omega)},$$

the complex transfer function can be written in the form

$$\mathbf{K}(i\omega) = e^{\Theta(i\omega)} = e^{A(\omega)} e^{i\varphi(\omega)} \tag{15.7}$$

The complex function $\Theta(i\omega)$ defined by expression (15.5) and characterizing the logarithmic attenuation of and also, the phase change in the two-port network can be called the *transmission constant* of this network, by analogy with the term used in the theory of long lines.

After transforming from the function $\mathbf{K}(i\omega)$ to the function $\Theta(i\omega)$, the problem is reduced to finding the relation between $A(\omega)$ and $\varphi(\omega)$, i.e., between the real and imaginary parts of the complex function $\Theta(i\omega)$.

For this purpose, let us use the following equation proved in the theory of functions of a complex variable:

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Theta(p)\, dp}{p - i\omega_1} = \frac{1}{2} \Theta(i\omega_1) + \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\Theta(i\omega)\, d(i\omega)}{i\omega - i\omega_1} \tag{15.8}$$

where ω_1 is a fixed arbitrary quantity.

The integration path on the left-hand side of this expression coincides with the axis $i\omega$ ($c \to 0$), the point $i\omega_1$ being passed from the right. The first term on the right-hand side is half the residue at the point $i\omega_1$, i.e., the integral along a semicircle of an infinitely

small radius r. In fact, on this circle the function $\Theta(p)$ is equal to $\Theta(i,\omega_1)$ to within the higher-order infinitesimals, therefore the denominator on the left-hand side of (15.8) can be written in the form

$$p - i\omega_1 = re^{i\psi}$$

where ψ is the argument of the vector drawn from the point $i\omega_1$ to the point lying on the circle of radius r, so that

$$dp = ire^{i\psi}d\psi$$

Therefore, the integral along the semicircle of radius r is

$$\frac{1}{2\pi i}\int_{i(\omega_1-r)}^{i(\omega_1+r)}\frac{\Theta(p)}{p-i\omega_1}dp = \frac{1}{2\pi i}\int_0^{\pi}\frac{\Theta(p)}{re^{i\psi}}ire^{i\psi}d\psi$$

$$= \frac{1}{2\pi i}\Theta(i\omega_1)\,i\pi = \frac{1}{2}\Theta(i\omega_1)$$

The second term on the right-hand side of (15.8) is the integral along the imaginary axis except for the singular point $i\omega_1$ (the principal value of the integral).

It should be noted that as the magnitude of p increases, and on condition that $\mathrm{Re}(p) > 0$, the function $\Theta(p)$ tends to zero. Therefore, the integral of the function $\Theta(p)/(p - i\omega_1)$ along an arc of an infinitely large radius R, lying in the right-hand half-plane, is equal

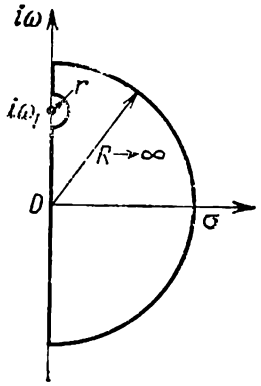


Fig. 15.1. Integration contour on the plane of complex variable

to zero. Hence, the integral on the left-hand side of equation (15.8) can be replaced by the contour integral shown in Fig. 15.1.

Let us impose the condition that the function $\Theta(p)$ *has no poles in the right-hand half-plane.* Then, according to the Cauchy theorem, the integral on the left-hand part of (15.8) is zero. Equating the left-hand side of equation (15.8) to zero and substituting (15.5) for $\Theta(i\omega)$, we obtain

$$\frac{1}{2}\left[A(\omega_1)+i\varphi(\omega_1)\right]+\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{A(\omega)\,d\omega}{\omega-\omega_1}+\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\varphi(\omega)\,d\omega}{\omega-\omega_1}=0$$

Equating to zero the real and imaginary parts of this expression, we get the following two equations:

$$A(\omega_1) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\varphi(\omega)}{\omega-\omega_1}d\omega$$

$$\varphi(\omega_1) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{A(\omega)}{\omega-\omega_1}d\omega$$

Thus, we have arrived at the Hilbert transforms (see Sec. 3.8). *The function* $A(\omega_1)$ *is the Hilbert transform (harmonic conjugate) of the function* $\varphi(\omega_1)$. Since $A(\omega)$ is an even and $\varphi(\omega)$ an odd function, we have

$$\int_{-\infty}^{\infty}\frac{A(\omega)}{\omega-\omega_1}d\omega = \int_0^{\infty}A(\omega)\left(\frac{1}{\omega-\omega_1}+\frac{1}{-\omega-\omega_1}\right)d\omega$$

$$= 2\omega_1\int_0^{\infty}\frac{A(\omega)}{\omega^2-\omega_1^2}d\omega$$

$$\int_{-\infty}^{\infty}\frac{\varphi(\omega)}{\omega-\omega_1}d\omega = \int_0^{\infty}\varphi(\omega)\left(\frac{1}{\omega-\omega_1}-\frac{1}{-\omega-\omega_1}\right)d\omega = 2\int_0^{\infty}\frac{\omega\varphi(\omega)}{\omega^2-\omega_1^2}d\omega$$

Substituting these results into the preceding expressions, we come to the following final formulas:

$$\varphi(\omega_1) = \frac{2\omega_1}{\pi}\int_0^{\infty}\frac{A(\omega)}{\omega^2-\omega_1^2}d\omega \qquad (15.9)$$

$$A(\omega_1) = -\frac{2}{\pi}\int_0^{\infty}\frac{\omega\varphi(\omega)}{\omega^2-\omega_1^2}d\omega \qquad (15.10)$$

Thus, the phase characteristic $\varphi(\omega_1)$ at some fixed frequency ω_1 is expressed through the logarithmic attenuation $A(\omega)$, the latter being integrated between the limits $\omega = 0$ and $\omega = \infty$. A similar rule applies to the logarithmic attenuation as well. Thus, to define one of the characteristics at some frequency, one has to know the change of the other within the entire frequency range.

Replacing in (15.9) $A(\omega)$ by $K(\omega)$ from formula (15.6), we obtain the sought for dependence between the phase- and amplitude-frequency characteristics:

$$\varphi(\omega_1) = \frac{2\omega_1}{\pi}\int_0^{\infty}\frac{\ln K(\omega)}{\omega^2-\omega_1^2}d\omega \qquad (15.11)$$

The above condition as to the absence of poles of the function $\Theta(p)$ in the right-hand half-plane is equivalent to the requirement that there should be no *poles and zeros* of the function $\mathbf{K}(p)$ in the same half-plane (since at the points of the plane p, where $\mathbf{K}(p)$ is zero, $\ln K(\omega)$ goes to $-\infty$).

Therefore, we can phrase the following important statement: *the one-to-one correspondence between the amplitude-frequency and phase characteristics is typical only of two-port networks whose transfer function* $\mathbf{K}(p)$ *has no zeros in the right-hand half-plane* p.

The two-port networks satisfying this condition are called *minimum-phase*. These include ordinary oscillatory circuits, filters, and circuits with no cross coupling. *Nonminimum-phase* networks include bridge circuits and some other special circuits.

Thus, if it is required that the two-port network being synthesized should be minimum-phase, its transfer function $\mathbf{K}(p)$ must have no zeros in the right-hand half-plane p. In this case, both polynomials $P(p)$ and $Q(p)$ in expression (15.1) must be the Hurwitz polynomials.

Let us dwell upon some properties of the amplitude-frequency and phase characteristics of the minimum-phase two-port networks. From expressions (15.9) and (15.10) it follows that the values of the integrals entering into them are determined by the character of change of $A(\omega)$ and $\varphi(\omega)$ in the vicinity of ω_1, since the absolute value of the fraction $1/(\omega^2 - \omega_1^2)$ rapidly decreases as ω moves away from ω_1. It should be noted that the integral of this fraction (whose sign changes at the point $\omega = \omega_1$) taken between 0 and ∞ is equal to zero*. Therefore, if we assume that for a certain physical circuit the attenuation $A(\omega) = A_0$, i.e., it is constant for all frequencies from 0 to ∞, then

$$\varphi(\omega_1) = \frac{2\omega_1}{\pi} \int_0^\infty \frac{A_0}{\omega^2 - \omega_1^2}\, d\omega = 0$$

Consequently, the logarithmic attenuation (and hence, the amplitude-frequency characteristic) can be made uniform over the entire range only for a circuit with zero phase characteristic, i.e., if the circuit consists of purely ohmic resistances.

On the other hand, the constant A_0 added to the attenuation $A(\omega)$ does not change the phase characteristic, so that expression (15.9) can be written in a more general form

$$\varphi(\omega_1) = \frac{2\omega_1}{\pi} \int_0^\infty \frac{A(\omega) - A_0}{\omega^2 - \omega_1^2}\, d\omega$$

* That is, the principal value of the integral is equal to zero everywhere except the singular point.

Physically, this means only a change in the scale of the amplitude-frequency characteristic, e.g., by means of an amplifier having a uniform amplitude-frequency characteristic or a voltage divider (potentiometer) made up of purely ohmic resistances. (In the first case A_0 must be taken positive, in the second, negative.)

It can also be shown that if the function $A(\omega)$ changes but slightly near the frequency ω_1, the phase characteristic defined by expression (15.9) changes linearly, while in the sub-bands characterized by comparatively fast variations of $A(\omega)$ the phase characteristic $\varphi(\omega)$ changes nonlinearly. In other words, linear sections of the phase-frequency characteristic correspond to linear sections of the amplitude-frequency characteristic.

In addition, where $K(\omega)$ passes through its maximum, i.e., within the passband of the circuit, the slope of the phase characteristic is negative $\left[\frac{d\varphi(\omega)}{d\omega} < 0\right]$. Correspondingly, where $K(\omega)$ passes through the minimum (outside the passband), the slope of the phase characteristic is positive $\left[\frac{d\varphi(\omega)}{d\omega} > 0\right]$. This is well illustrated by the amplitude-frequency and phase characteristics of the oscillatory circuits discussed in Sec. 5.7 (e.g., the resonance and phase characteristics in Fig. 5.18).

15.4. REPRESENTATION OF A GENERAL TWO-PORT NETWORK BY CASCADE CONNECTION OF ELEMENTARY TWO-PORT NETWORKS

With the zeros and poles being given, it is advisable to represent the transfer function $\mathbf{K}(p)$ in the form of a product of factors each of which can be a transfer function of a simple elementary two-port network. For example, let the transfer function

$$\mathbf{K}(p) = \frac{a_0}{b_0} \frac{p}{(p-p_{p1})(p-p_{p2})(p-p_{p3})} \tag{15.12}$$

of the two-port network being synthesized have a zero at the point $p = 0$ and three poles, one of which, at the point $p_{p1} < 0$, being real and the other two, p_{p2} and $p_{p3} = p_{p2}^*$, complex.

Taking into account the equation

$$(p-p_{p2})(p-p_{p2}^*) = p^2 - 2\,\mathrm{Re}(p_{p2})\,p + |p_{p2}|^2 \tag{15.13}$$

we write (15.12) in the form

$$\mathbf{K}(p) = \frac{a_0}{b_0} \frac{p}{(p-p_{p1})} \frac{1}{p^2 - 2\mathrm{Re}(p_{p2})\,p + |p_{p2}|^2} = \frac{a_0}{b_0} \mathbf{K}_1(p)\,\mathbf{K}_2(p) \tag{15.14}$$

The transfer function $\mathbf{K}_1(p)$ is realized by a first-order link (R-C or R-L circuit). In fact, for an R-C circuit (Fig. 6.6*a*) where the voltage

is taken from across the resistor, the transfer function is

$$\mathbf{K}_1(p) = \frac{R}{R + 1/Cp} = \frac{p}{p + 1/RC} \tag{15.15}$$

from which it follows that $p_{p1} = -1/RC$, $RC = -1/p_{p1}$.

When an R-L circuit is used (Fig. 6.7a), we have

$$\mathbf{K}_1(p) = \frac{Lp}{R + Lp} = \frac{p}{p + R/L} \tag{15.15'}$$

whence $p_{p1} = -R/L$.

The function $\mathbf{K}_2(p)$ is realized by a second-order link. This question is discussed in detail in the next section.

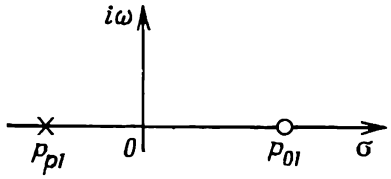


Fig. 15.2. Location of a zero and a pole on the p-plane for a nonminimum-phase two-port network

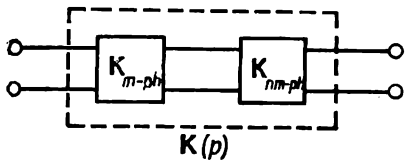


Fig. 15.3. Equivalent circuit of a two-port network with the zero on the right-hand half-plane

When synthesizing a nonminimum-phase two-port network, the division of its transfer function into simple factors has some specific features. Let us illustrate them with the example of the function

$$\mathbf{K}(p) = B\frac{p - p_{01}}{p - p_{p1}} \tag{15.16}$$

where p_{01} is a real positive quantity, while p_{p1} is a real but negative quantity (Fig. 15.2).

Let us multiply the numerator and denominator in (15.16) by $p + p_{01}$. Then

$$\mathbf{K}(p) = B\frac{p + p_{01}}{p - p_{p1}}\frac{p - p_{01}}{p + p_{01}} = \mathbf{K}_{m\text{-}ph}(p)\,\mathbf{K}_{nm\text{-}ph}(p) \tag{15.17}$$

where

$$\mathbf{K}_{m\text{-}ph}(p) = B\frac{p + p_{01}}{p - p_{p1}} \tag{15.18}$$

stands for the transfer function of a minimum-phase two-port network (since the zero is located at the point $p = -p_{01}$, i.e., in the left-hand half-plane) and

$$\mathbf{K}_{nm\text{-}ph}(p) = \frac{p - p_{01}}{p + p_{01}} \tag{15.19}$$

stands for the transfer function of a nonminimum-phase two-port network.

The right-hand part of expression (15.17) corresponds to the transfer function of two cascade-connected two-port networks. Consequently, the given two-port network with the transfer function $\mathbf{K}(p)$ defined by expression (15.16) can be replaced by an equivalent cascade connection of two two-port networks $\mathbf{K}_{m\text{-}ph}(p)$ and $\mathbf{K}_{nm\text{-}ph}(p)$ (Fig. 15.3).

Let us consider the second two-port network $\mathbf{K}_{nm\text{-}ph}$ in more detail. Transforming from p to $i\omega$, we write this function as

$$\mathbf{K}_{nm\text{-}ph}(i\omega) = \frac{i\omega - p_{01}}{i\omega + p_{01}} \tag{15.20}$$

Since p_{01} is a real number, the modulus of this function is equal to unity at all frequencies from $\omega = 0$ to ∞, while the argument is

$$\varphi(\omega) = \arg(i\omega - p_{01}) - \arg(i\omega + p_{01})$$

$$= -\operatorname{arc\,tan}\frac{\omega}{p_{01}} - \operatorname{arc\,tan}\frac{\omega}{p_{01}} = -2\operatorname{arc\,tan}\frac{\omega}{p_{01}}$$

Thus, the transfer function

$$\mathbf{K}_{nm\text{-}ph}(i\omega) = e^{-i2\operatorname{arc\,tan}\omega/p_{01}} \tag{15.21}$$

Thus, the two-port network with the transfer function of the form (15.20) must let pass through (uniformly) all frequencies from 0 to ∞. Such two-port networks as make it possible to correct the phase-frequency characteristic without changing the amplitude-frequency characteristic are called phase-correcting circuits. The realization of such circuits is discussed in Sec. 15.6.

The treatment of expression (15.14) as the transfer function of a cascade connection of mutually independent two-port networks $\mathbf{K}_1(p)$ and $\mathbf{K}_2(p)$ enables one to reduce the synthesis of a complex two-port network to that of simple links. An increase in the number of zeros and poles in the transfer function results merely in a respective increase in the number of the links. Naturally, such an approach makes sense and is permissible only with an adequate decoupling of the elementary two-port networks. The use of emitter followers and some modern microelectronic devices ensures that this requirement is fulfilled. Where the mutual influence of the elementary two-port networks cannot be neglected, one has to resort to more complex synthesis methods which are discussed in special literature [7] through [9].

15.5. REALIZATION OF A TYPICAL SECOND-ORDER LINK

In accordance with (15.14), let us represent the transfer function of an elementary two-port network in the form

$$\mathbf{K}(p) = a_0 \frac{1}{p^2 - 2\operatorname{Re}(p_p)\,p + |p_p|^2} = a_0 \frac{1}{p^2 + b_1 p + b_2} \tag{15.22}$$

where the constant coefficients $b_1 = -2\mathrm{Re}(p_p)$ and $b_2 = |p_p|^2$.

First, let us consider the realization of the function $\mathbf{K}(p)$ by means of a circuit comprising an inductance coil L, a capacitor C, and a resistor R (Fig. 15.4). The resistance of the resistor, which is the load of the two-port network, is considered to be given. One of the elements of the circuit Z_{12}, Z_2 must be inductive and the other, capa-

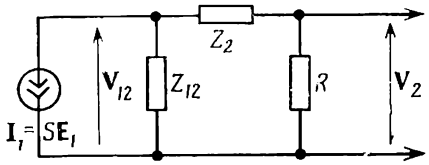


Fig. 15.4. Realization of a typical second-order link

citive. By the current source exciting the circuit one may imply, for example, the collector circuit of a common-emitter transistor amplifier (see Sec. 5.4, Fig. 5.12*b*). The internal conductance of the current source is neglected. The current $\mathbf{I}_1$ is equal to $S\mathbf{E}_1$, where $\mathbf{E}_1$ is the base-to-emitter voltage.

The voltage across the element Z_{12} can be defined by the expression

$$\mathbf{V}_{12} = \frac{Z_{12}(Z_2+R)\,S\mathbf{E}_1}{Z_{12}+Z_2+R}$$

and the voltage across the resistor R, by

$$\mathbf{V}_2 = \frac{R}{R+Z_2}\mathbf{V}_{12} = \frac{R}{(R+Z_2)}\,\frac{Z_{12}(Z_2+R)}{(Z_{12}+Z_2+R)}\,S\mathbf{E}_1$$

Hence,

$$\mathbf{K}(p) = \frac{\mathbf{V}_2}{\mathbf{E}_1} = S\,\frac{Z_{12}R}{Z_{12}+Z_2+R} \tag{15.23}$$

From comparison between this expression and (15.22) it is clear that to obtain a real numerator one should set $Z_{12} = 1/Cp$ and $Z_2 = Lp$. In this case

$$\mathbf{K}(p) = SR\,\frac{1}{Cp\left(\frac{1}{Cp}+Lp+R\right)} = \frac{SR}{LC}\,\frac{1}{\left(p^2+\frac{R}{L}p+\frac{1}{LC}\right)} \tag{15.24}$$

Comparison between (15.24) and (15.22) yields

$$R/L = b_1, \quad 1/LC = b_2, \quad SR/LC = a_0$$

whence

$$L = R/b_1, \quad C = 1/b_2L, \quad a_0 = SRb_2 \tag{15.25}$$

Thus, the circuit diagram of the sought for circuit assumes the form shown in Fig. 15.5*a*.

In a similar way, one can easily show that to the transfer function of the form

$$K(p) = a_0 \frac{p^2}{p^2 + b_1 p + b_2} \tag{15.26}$$

corresponds the circuit shown in Fig. 15.5*b*, whose parameters L and C are expressed through the coefficients b_1 and b_2 by the same

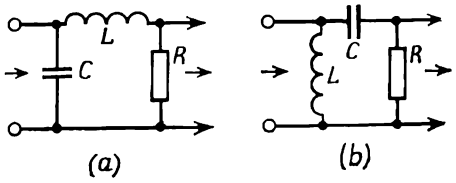


Fig. 15.5. Realization of a transfer function according to (*a*) expression (15.24) and (*b*) expression (15.26)

relations (15.25) as in the circuit shown in Fig. 15.5*a*, the only difference being in the constant coefficient $a_0 = SR_1$.

In integrated circuits that do not allow one to use inductance coils, the second-order circuit is realized by means of an active R-C circuit. One of possible versions of such a circuit is shown in Fig. 15.6*a*. The required properties here are obtained by using an operational amplifier K_0 and feedback. In this circuit, the amplification factor of the amplifier is fairly low (a few units). The basic requirements for the

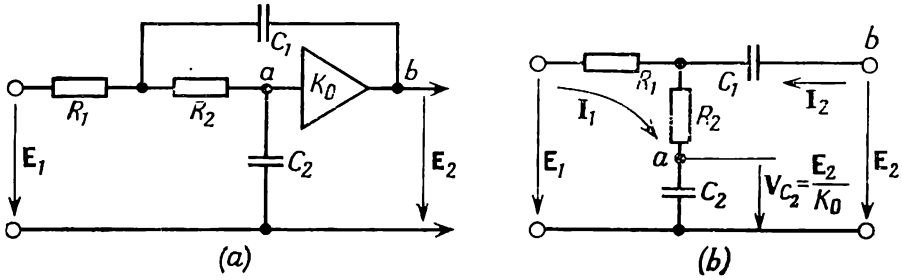


Fig. 15.6. (*a*) Active second-order R-C circuit and (*b*) its equivalent circuit

amplifier are a very high input resistance and a very low (close to zero) output resistance, and also, the absence of back reaction. If these requirements are met, the amplifier can be regarded as an ideal voltage source (voltage-controlled) and this makes it possible, when defining the currents and voltages in the circuit shown in Fig. 15.6*a*, to consider the points *a* and *b* to be open-circuited and take the output voltage to be equal to $K_0 V_{C_2}$, where V_{C_2} is the voltage across the capacitor C_2. These assumptions lead to the equivalent circuit of Fig. 15.6*b*, in which the amplifier K_0 is omitted and

its effect is taken into account by setting the relation between the voltage across the capacitor C_2 and the output voltage to be $\mathbf{V}_{C_2} = \mathbf{E}_2/K_0$.

Applying the general equations (5.4) of the two-port network to the circuit shown in Fig. 15.6*b* and taking into account the additional condition

$$\mathbf{E}_2 = K_0 (\mathbf{I}_1 + \mathbf{I}_2)/C_2 p$$

we obtain

$$\left.\begin{aligned} \mathbf{E}_1 &= Z_{11}\mathbf{I}_1 + Z_{12}\mathbf{I}_2 \\ \mathbf{E}_2 &= Z_{21}\mathbf{I}_1 + Z_{22}\mathbf{I}_2 = K_0 (\mathbf{I}_1 + \mathbf{I}_2)/C_2 p \end{aligned}\right\} \qquad (15.27)$$

Here

$$Z_{11} = R_1 + R_2 + 1/C_2 p, \quad Z_{12} = R_2 + 1/C_2 p$$
$$Z_{21} = R_2 + 1/C_2 p, \quad Z_{22} = R_2 + 1/C_1 p + 1/C_2 p$$

Eliminating the current $\mathbf{I}_2$ from the first equation of (15.27), we obtain, after simple manipulations, the following expression for the transfer function of the two-port network:

$$\mathbf{K}(p) = \frac{\mathbf{E}_2}{\mathbf{E}_1} = \frac{\dfrac{K_0}{R_2 C_2}}{C_1 R_1 p^2 + \left(1 + \dfrac{R_1}{R_2} + \dfrac{C_1 R_1}{C_2 R_2} - K_0 \dfrac{C_1 R_1}{C_2 R_2}\right) p + \dfrac{1}{C_2 R_2}}$$

$$= \frac{K_0/C_1 C_2 R_1 R_2}{p^2 + (1/C_1 R_1 + 1/C_1 R_2 + 1/C_2 R_2 - K_0/C_2 R_2)\, p + 1/C_1 C_2 R_1 R_2} \qquad (15.28)$$

Further, the problem of synthesis is reduced to the selection of resistors, capacitors, and amplification factor K_0 providing for the required values of the coefficients b_1 and b_2 of polynomial (15.22):

$$b_1 = \left(\frac{1}{R_1 C_1} + \frac{1}{R_2 C_1} + \frac{1}{R_2 C_2} - \frac{K_0}{R_2 C_2}\right), \quad b_2 = \frac{1}{R_1 R_2 C_1 C_2} \qquad (15.29)$$

From the first equation we can obtain the following expression for the required amplification factor:

$$K_0 = 1 + C_2/C_1 + R_2 C_2/R_1 C_1 - b_1 R_2 C_2 \qquad (15.30)$$

The relations thus obtained will be illustrated in Sec. 15.8.

15.6. REALIZATION OF A PHASE-CORRECTING CIRCUIT

Nonminimum-phase two-port networks with the amplitude-frequency characteristic uniform in the range $0 < |\omega| < \infty$ are realized by means of bridge circuits. A simple circuit of this type, composed of capacitors and resistors, was described in Sec. 14.2. The disadvantage of this circuit is that it requires infinitely large resistance of the load connected across the bridge diagonal 2—2′ (see Fig. 14.5). A bridge circuit consisting of inductance coils and capacitors is free from this disadvantage, provided that the load resistance R_l is

selected appropriately. In the circuit diagrams shown in Fig. 15.7*a* and *b* and differing only in configuration, the admittance Y_l is equal to $1/R_l$, while the admittances Y_a and Y_b of the bridge branches are selected from the condition

$$Y_a Y_b = Y_l^2 = 1/R_l^2$$

To define the transfer function of the two-port network under consideration, let us use the equivalent circuit shown in Fig. 5.4*a*. In

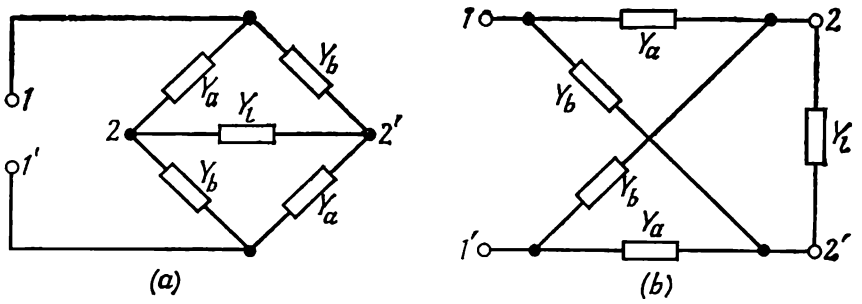


Fig 15.7 Symmetrical bridge two-port network of different configurations

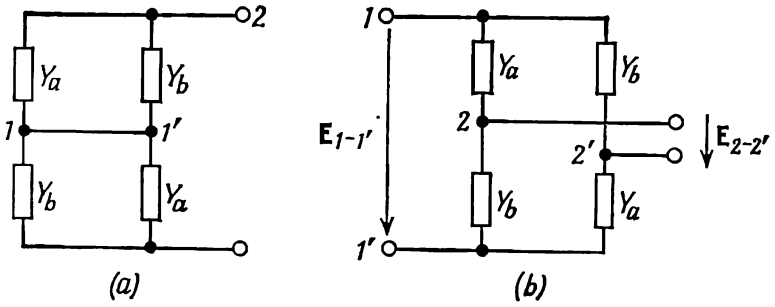


Fig 15.8. Equivalent circuit of a bridge two-port network with (*a*) short-circuited input and (*b*) open-circuited input

this case, according to equation (5.17),

$$K_E = \frac{E_{2\text{-}2'}}{E_{1\text{-}1'}} = \frac{-Y_{21}}{Y_{22}+Y_l} \qquad (15.31)$$

The admittance Y_{22} is found on condition that the terminals 1 and 1′ (Fig. 15.8*a*) are short-circuited. It is clear that

$$Y_{22} = \frac{1}{2}(Y_a + Y_b) \qquad (15.32)$$

Since the circuit is symmetrical, we can write

$$Y_{11} = Y_{22} = \frac{1}{2}(Y_a + Y_b) \qquad (15.33)$$

The admittance $Y_{21} = Y_{12}$ is easily determined from the expression for the transfer function under no-load conditions ($Y_l = 0$, Fig. 15.8*b*):

$$K_{E_{xx}} = \frac{1/Y_b}{1/Y_a + 1/Y_b} - \frac{1/Y_a}{1/Y_a + 1/Y_b} = \frac{Y_a - Y_b}{Y_b + Y_a} = -\frac{Y_{21}}{Y_{22}}$$

Substituting Y_{22} by expression (15.32), we get

$$Y_{21} = Y_{12} = \frac{1}{2}(Y_b - Y_a) \tag{15.34}$$

On the basis of (15.32) and (15.34), expression (15.31) is reduced to the form

$$K_E = \frac{Y_a - Y_b}{Y_a + Y_a + 2Y_l} \tag{15.35}$$

and taking into account the equality $Y_b = Y_l^2/Y_a$

$$K_E = \frac{Y_a - Y_l^2/Y_a}{Y_a + Y_l^2/Y_a + 2Y_l} = \frac{Y_a - Y_l}{Y_a + Y_l} \tag{15.36}$$

Transforming from admittances to impedance, we obtain

$$K_E = \frac{1/Z_a - 1/R_l}{1/Z_a + 1/R_l} = \frac{R_l - Z_a}{R_l + Z_a} \tag{15.37}$$

If the impedance Z_a is formed by a capacitance and Z_b, by an in-

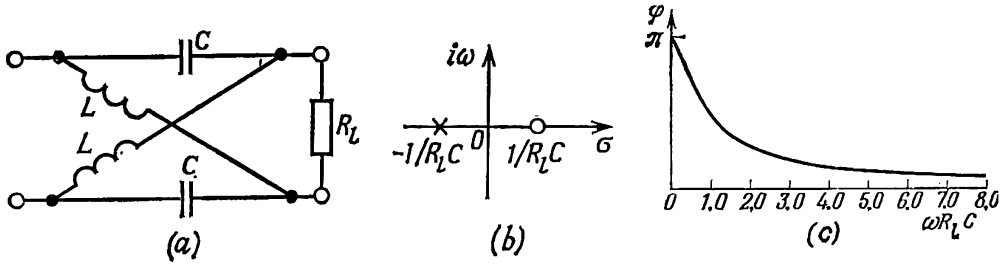


Fig. 15.9. Bridge two-port network

(*a*) circuit diagram; (*b*) location of the zero and the pole on the *p*-plane; (*c*) phase-frequency characteristic

ductance (Fig. 15.9*a*), $Z_a = 1/Cp$ and $Z_b = Lp$. In this case

$$K_E(p) = \frac{R_l - 1/Cp}{R_l + 1/Cp} = \frac{p - 1/R_lC}{p + 1/R_lC} \tag{15.38}$$

Thus, the zero of the transfer function is $p_0 = 1/R_lC$ and the pole, $p_p = -1/R_lC$ (Fig. 15.9*b*).

Transforming from p to $i\omega$, we obtain

$$K_E(i\omega) = \frac{i\omega - 1/R_lC}{i\omega + 1/R_lC} = e^{i\varphi(\omega)} \tag{15.39}$$

The amplitude-frequency characteristic $K_E(\omega) = 1$, while the phase-frequency characteristic (Sig. 15.9*c*) is

$$\varphi(\omega) = \pi - 2 \arctan \omega R_l C \qquad (15.40)$$

Thus, if the zero $p_0 > 0$ of the transfer function and the load resistance R_l are specified, the elements of the bridge network (Fig. 15.9*a*) are defined by the relations

$$C = 1/R_l p_0, \quad L = R_l^2 C \qquad (15.41)$$

The input impedance (or admittance) is an important parameter of a two-port network, particularly where it is used in a cascade circuit. Setting up an expression for the input admittance, similar

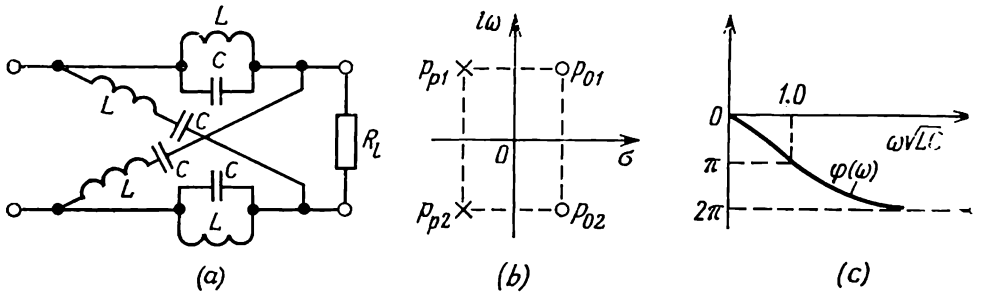


Fig. 15.10. Bridge two-port network

(*a*) circuit diagram; (*b*) location of the zeros and poles on the *p*-plane; (*c*) phase-frequency characteristic

to (5.23), and substituting therein Y_{11} and $Y_{21} = Y_{12}$ according to formulas (15.33) and (15.34), we make certain that

$$Y_{in} = Y_l = 1/R_l \qquad (15.42)$$

Thus, the input admittance of a bridge two-port network with a matched load ($R_l = \sqrt{L/C}$) is equal to $1/R_l$ regardless of frequency.

A nonminimum-phase two-port network with two complex-conjugate zeros in the right-hand half-plane p can be realized using the bridge circuit shown in Fig. 15.10*a*.

In this circuit

$$Z_a = \frac{Lp\,(1/Cp)}{Lp + 1/Cp} = \frac{Lp}{LCp^2 + 1}$$

$$Z_b = Lp + \frac{1}{Cp} = \frac{LCp^2 + 1}{Cp}, \qquad Z_a Z_b = \frac{L}{C} = R_l^2$$

The transfer function

$$K_E(p) = \frac{R_l - Z_a}{R_l + Z_a} = \frac{p^2 - p/R_l C + 1/LC}{p^2 + p/R_l C + 1/LC} = \frac{(p - p_{01})(p - p_{02})}{(p - p_{p1})(p - p_{p2})}$$

where the zeros are

$$p_{01,2} = \frac{1}{2R_lC} \pm \sqrt{\left(\frac{1}{2R_lC}\right)^2 - \frac{1}{LC}}$$

$$= \frac{1}{2R_lC} \pm i\sqrt{\frac{1}{LC} - \left(\frac{1}{2R_lC}\right)^2} = \frac{1}{2R_lC} \pm i\omega_0$$

and the poles are

$$p_{p1,2} = -\frac{1}{2R_lC} \pm i\sqrt{\frac{1}{LC} - \left(\frac{1}{2R_lC}\right)^2} = -\frac{1}{2R_lC} \pm i\omega_0$$

The poles and zeros of the transfer function are located symmetrically with respect to the axis $i\omega$ (Fig. 15.10*b*).

On the frequency axis, the transfer function is

$$\mathbf{K}_E(i\omega) = \frac{-\omega^2 + 1/LC - i\omega/R_lC}{-\omega^2 + 1/LC + i\omega/R_lC} \tag{15.43}$$

As in the preceding circuit, $K_E(\omega) = 1$. The phase-frequency characteristic (Fig. 15.10*c*) is easily reduced to the form

$$\varphi(\omega) = -2\arctan\frac{\omega\sqrt{LC}}{1-\omega^2LC} \tag{15.44}$$

The bridge two-port networks have found wide application in the synthesis of circuits with amplitude-frequency and phase-frequency characteristics which cannot be realized by means of minimum-phase two-port networks.

Nonminimum-phase circuits are also used in the synthesis of networks matched to a specified signal, since one-to-one relation does not necessarily exist between the modulus and the argument of the spectral density of the signal. For example, the matched filters discussed in Ch. 12 used delay lines with transfer functions of the form $e^{i\omega t} = 1 \times e^{-i\varphi(\omega)}$. However, as mentioned in Sec. 15.1, the classical theory of synthesis of electric circuits is inapplicable to such circuits.

15.7. SYNTHESIS OF A TWO-PORT NETWORK BY THE GIVEN AMPLITUDE-FREQUENCY CHARACTERISTIC

When synthesizing low-pass filters, high-pass filters, band filters, etc., no special requirements are usually placed upon their phase characteristics. It is assumed that making the amplitude-frequency characteristic of a minimum-phase two-port network satisfactorily uniform in a given frequency band also provides for the linearity of the phase-frequency characteristic in this band.

Using general expression (15.1), let us represent the complex transfer function $\mathbf{K}(i\omega)$ in the form

$$\mathbf{K}(i\omega) = \mathbf{K}(p)_{p=i\omega} = \frac{P(p)}{Q(p)}\Big|_{p=i\omega} \tag{15.45}$$

and then proceed to the square of the modulus

$$K^2(\omega) = \mathbf{K}(i\omega)\,\mathbf{K}(-i\omega) = \frac{P(p)\,P(-p)}{Q(p)\,Q(-p)}\Big|_{p=i\omega} \qquad (15.46)$$

thus excluding from consideration the phase-frequency characteristic of the two-port network.

The modulus of the transfer function, which is even with respect to frequency, can be regarded as a function of ω^2. The same applies

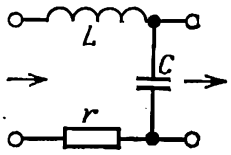


Fig. 15.11. Simple two-port network with two poles

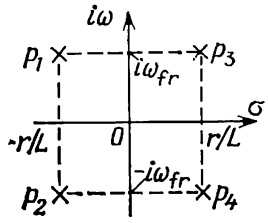


Fig. 15.12. Quadrantal symmetry of poles

to the moduli $|P(i\omega)|$ and $|Q(i\omega)|$. Therefore, expression (15.46) can be written in the form

$$|\mathbf{K}(i\omega)|^2 = \mathbf{K}(p)\,\mathbf{K}(-p)\Big|_{p=i\omega} = \frac{A(\omega^2)}{B(\omega^2)} = \frac{A(-p^2)}{B(-p^2)}\Big|_{p=i\omega} \qquad (15.47)$$

where

$$A(-p^2) = P(p)\,P(-p), \qquad B = (-p^2) = Q(p)\,Q(-p)$$

Transforming from the imaginary axis $i\omega$ to any point of the p-plane, we obtain the following expression

$$\mathbf{K}(p)\,\mathbf{K}(-p) = A(-p^2)/B(-p^2) \qquad (15.48)$$

The poles and zeros of the function $A(-p^2)/B(-p^2)$ show quadrantal symmetry: for each complex-conjugate pair in the left-hand half-plane p there is a mirror-image pair in the right-hand half-plane.

Let us illustrate this assertion with an example of a simple two-port network (Fig. 15.11) whose transfer function is given by

$$\mathbf{K}(i\omega) = \frac{1/i\omega C}{i\omega L + r + 1/i\omega C}, \qquad \mathbf{K}(p) = \frac{1/LC}{p^2 + pr/L + 1/LC}$$

For the complex-conjugate function $\mathbf{K}(-i\omega)$, the following relations hold:

$$\mathbf{K}(-i\omega) = \frac{1/(-i\omega C)}{-i\omega L + r + 1/(-i\omega C)}, \qquad \mathbf{K}(-p) = \frac{1/LC}{p^2 - pr/L + 1/LC}$$

Consequently,

$$|\mathbf{K}(i\omega)|^2 = \frac{(1/LC)^2}{(\omega^2 - 1/LC)^2 + (r/L\omega)^2} = \frac{(1/LC)^2}{B(-p^2)}\Big|_{p=i\omega}$$

$$\mathbf{K}(p)\,\mathbf{K}(-p) = \frac{(1/LC)^2}{(p^2 + pr/L + 1/LC)(p^2 - pr/L + 1/LC)} = \frac{(1/LC)^2}{B(-p^2)}$$

The poles of the function $(1/LC)^2/B(-p^2)$, which are the roots of the equation $B(-p^2) = 0$, are located at the points (Fig. 15.12)*

$$p_{1,2} = -r/2L \pm i\sqrt{1/LC - (r/2L)^2} = -\alpha \pm i\omega_{fr}$$

$$p_{3,4} = +r/2L \pm i\sqrt{1/LC - (r/2L)^2} = +\alpha \pm i\omega_{fr}$$

The poles of the transfer function $\mathbf{K}(p)$ in the given example, p_1 and p_2 are located only in the left-hand half-plane p. The same applies to the zeros of the transfer function, i.e., to the roots of the equation $A(-p^2) = 0$ (in the given example, there are no zeros) if the transfer function $\mathbf{K}(p)$ corresponds to a minimum-phase circuit, otherwise the zeros may be located in the right-hand half-plane p as well.

It should be also noted that the poles located on the imaginary axis can be only multiple (of multiplicity 2). Half of them must be related to $\mathbf{K}(p)$, the other half relating to $\mathbf{K}(-p)$.

From the above-listed properties of the function $K^2(\omega)$ it follows that the approximation of a given amplitude-frequency characteristic of a two-port network can be performed by using functions depending on ω^2 and, after transforming to the variable $p = \sigma + i\omega$, by using such functions as correspond to the above location of the poles and zeros on the p-plane.

15.8. SYNTHESIS OF A LOW-PASS FILTER. BUTTERWORTH FILTER

The amplitude-frequency characteristic of an ideal low-pass filter is shown in Fig. 15.13. When approximating the amplitude-frequency characteristic, the dimensionless (normalized) frequency $x = \omega/\omega_{cut}$, where ω_{cut} is the cutoff frequency, is plotted on the abscissa axis, while the normalized value $K(\omega/\omega_{cut}) = K(x)$ is plotted on the ordinate axis.

The approximating function for the ideal characteristic of the low-pass filter shown in Fig. 15.13 is given in the form

$$K(x) = 1/\sqrt{1 + F^2(x)} \qquad (15.49)$$

subject to the condition that the magnitude of the function $F(x)$ is minimum in the band $0 < x \leqslant 1$ and maximum for $x > 1$.

---

* Here and later in the text the subscript "p" in the pole designation "p_{pk}" is omitted.

A simplest function satisfying this condition is $F(x) = x^n = (\omega/\omega_{cut})^n$. In this case

$$K^2(x) = |K(ix)|^2 = 1/[1 + F^2(x)] = 1/(1 + x^{2n}) \quad (15.50)$$

The graphs of function (15.49) for several values of n are shown in Fig. 15.14. Approximating function (15.50) is known as the Butterworth function, and the filters synthesized on the basis of this function are called the Butterworth filters. For the cutoff frequency

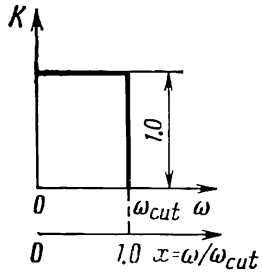


Fig. 15.13. Amplitude-frequency characteristic of an ideal low-pass filter

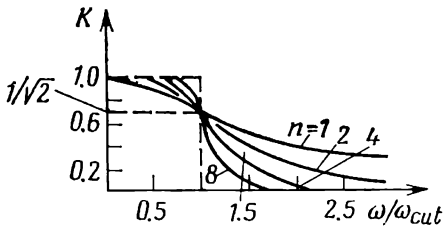


Fig. 15.14. Butterworth approximating functions

$x = 1$ ($\omega = \omega_{cut}$), the Butterworth functions of any order n are equal to 1/2 and this corresponds to the attenuation of the amplitude-frequency characteristic to $1/\sqrt{2}$ (3 dB). The Butterworth approximation is often called the maximum flat approximation.

When $K(x)$ is calculated in decibels, expression (15.49) is reduced to

$$K(x)_{dB} = 20 \log_{10} K(x) = -10 \log_{10} [1 + F^2(x)]$$
$$= -10 \log_{10} (1 + x^{2n})$$

If the dimensionless frequency x is represented in the form of a power of 2, i.e., $x = 2^y$, where y is the number of octaves, we obtain

$$K(x)_{dB} = -10 \log_{10} (1 + x^{2n}) = -10 \log_{10} (1 + 2^{2ny}) \quad (15.51)$$

The graph of K (in decibels) versus y is shown in Fig. 15.15. At the cutoff frequency ($x = 1$, $y = 0$) the attenuation is equal to 3 dB regardless of the order of n.

Outside the filter passband, for $x^{2n} \gg 1$ ($y > 1$), expression (15.51) defines a straight line

$$(1/K)_{dB} \approx 10 \log_{10} 2^{2ny} = 20ny \log_{10} 2 = 6ny \quad (15.52)$$

Thus, the attenuation of the amplitude-frequency characteristic is equal to $6n$ dB per octave (i.e., with the frequency x changing twofold and y changing by unity).

A satisfactory approximation of a rectangular characteristic (Fig. 15.13) by means of the Butterworth functions requires rather

high values of n. For example, if it is necessary that at $\omega = 3\omega_{cut}$ ($x = 3$) the attenuation of the amplitude-frequency characteristic be at least 40 dB, then $n \geqslant 40/6y$. In this case, $y = \log_{10}x/\log_{10}2 = 1.58$, hence $n \geqslant 4.25$, i.e., it is necessary that $n = 5$.

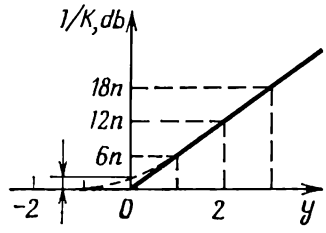


Fig. 15.15. Attenuation in a Butterworth filter as a function of the number of octaves

The next step after determining the order n of the filter is finding the poles of the transfer function. For this purpose, let us express (15.50) in the form of (15.47) by setting $ix = p$, $x^2 = -p^2$, and $x^{2n} = (-1)^n p^{2n}$ in (15.50):

$$|\mathbf{K}(ix)|^2 = \mathbf{K}(p)\,\mathbf{K}(-p)\,|_{p=ix} = \frac{1}{1+(-1)^n p^{2n}}\Big|_{p=ix} = \frac{1}{B(-p^2)}\Big|_{p=ix}$$

Now, considering the behaviour of the function $1/B(-p^2)$ on the p-plane, we find the poles as the roots of the equation

$$B(-p^2) = 1 + (-1)^n p^{2n} = 0$$

or

$$p^{2n} = -1/(-1)^n = (-1)^{-(n-1)} \tag{15.53}$$

Using the relations

$$(-1) = e^{-i\pi}, \quad (-1)^{n-1} = e^{-i\pi(n-1)}e^{-i2k\pi}$$

where k is any integer, we obtain for the kth root of equation (15.53) the following expression:

$$p_k = e^{i\pi[(n-1)+2k]/2n} \tag{15.54}$$

the number of roots being equal to the degree of equation (15.53).

The moduli of all the poles p_k are equal to unity, while the arguments are

$$\varphi_k = \pi\,[n + (2k - 1)]/2n \tag{15.55}$$

the difference between the arguments of any two adjacent roots being equal to π/n. Consequently, all the poles of the function $1/B(-p^2)$ lie on a unit-radius circle and divide this circle into equal arcs of π/n. The argument of the first pole is $\varphi_1 = \pi(n + 1)/2n$ and that of the last one is $\varphi_{2n} = \pi(5n - 1)/2n$.

The arrangement of the poles on the unit-radius circle for the Butterworth filters of the third ($n = 3$) and fourth ($n = 4$) orders is shown in Fig. 15.16.

In accordance with Sec. 15.7, the transfer function of the filter to be synthesized has its poles lying only in the left-hand half-plane. These poles are

$$p_k = -\sin\left(\frac{2k-1}{2n}\pi\right) + i\cos\left(\frac{2k-1}{2n}\pi\right), \quad k = 1, 2, \ldots, n \qquad (15.56)$$

It should be borne in mind that formulas (15.53) through (15.57) define the values of *normalized* variables p, i.e.,

$$p = (\sigma + i\omega)/\omega_{cut} \qquad (15.57)$$

All the poles form complex-conjugate pairs except for the only pole on the real axis, that corresponds to an odd n. For this pole, $k =$

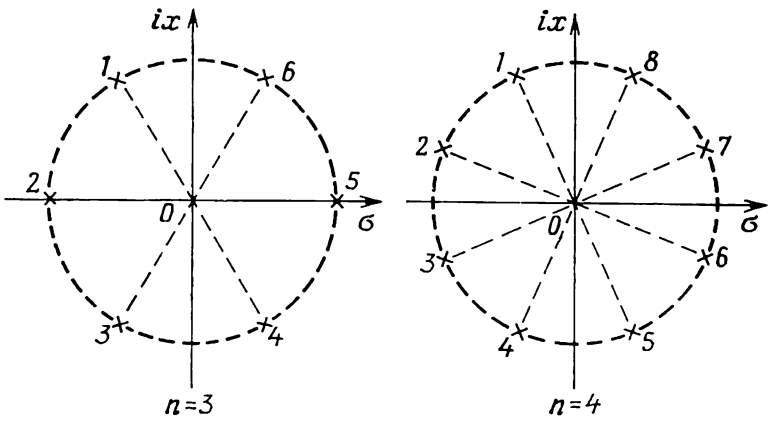


Fig. 15.16. Arrangement of the poles of the transfer function of a third- and a fourth-order Butterworth filter

$= (n + 1)/2$. Substituting p_k by formula (15.54) into general expression (15.2), we get the transfer function of the Butterworth filter. Let us give this expression for $n = 2$, 3 and 4.

For $n = 2$, the poles are $p_{1,2} = -1/\sqrt{2} \pm i/\sqrt{2}$, and by formulas (15.13) and (15.22) we find

$$\mathrm{K}(p) = \frac{1}{(p-p_1)(p-p_2)} = \frac{1}{p^2+\sqrt{2}\,p+1} = \frac{1}{p^2+b_1p+b_2} \qquad (15.58)$$

For $n = 3$, the poles are $p_1 = -0.5 + i\frac{\sqrt{3}}{2}$, $p_2 = -1$, and $p_3 =$ $= -0.5 - \mathrm{i}\frac{\sqrt{3}}{2} = p_1$.

The transfer function

$$\mathrm{K}(p) = \frac{1}{(p-p_1)(p-p_2)(p-p_3)} = \frac{1}{(p-p_2)} \frac{1}{(p-p_1)(p-p_1^*)}$$

$$= \frac{1}{(p+1)} \frac{1}{(p^2+p+1)} = \frac{1}{p^3+b_1p^2+b_2p+b_3} \qquad (15.59)$$

For $n = 4$, the transfer function is reduced to the form

$$\mathbf{K}(p) = \frac{1}{(p^2+0.765p+1)} \frac{1}{(p^2+1.848p+1)} = \frac{1}{p^4+b_1p^3+b_2p^2+b_3p+b_4} \quad (15.60)$$

The coefficients of the polynomials in the denominator of the Butterworth transfer functions are given in many handbooks on filter calculations [8], [12].

The last step of the synthesis of the low-pass filter is the selection of the elements for typical second-order links (with odd n, additionally for one first-order link).

Let us give an example of synthesis of a second-order Butterworth filter ($n = 2$) comprising a single link whose transfer function is

$$\mathbf{K}(p) = \frac{1}{p^2+\sqrt{2}\,p+1}, \qquad b_1 = \sqrt{2},\ b_2 = 1$$

Transforming in expression (15.28) to a normalized frequency variable $p = (\sigma + i\omega)/\omega_{cut}$ [as in (15.58)], we reduce it to the form

$$\mathbf{K}(p) = \frac{\frac{K_0}{R_1R_2C_1C_2}}{\omega_{cut}^2p^2+\frac{\omega_{cut}}{R_2C_2}\left(\frac{R_2C_2}{R_1C_1}+\frac{C_2}{C_1}+1-K_0\right)p+\frac{1}{R_1R_2C_1C_2}} = \frac{\frac{K_0}{\omega_{cut}^2R_1R_2C_1C_2}}{p^2+\frac{1}{\omega_{cut}R_2C_2}\left(\frac{R_2C_2}{R_1C_1}+\frac{C_2}{C_1}+1-K_0\right)p+\frac{1}{\omega_{cut}^2R_1R_2C_1C_2}} \quad (15.61)$$

Equating the denominators in expressions (15.58) and (15.61), we obtain the following conditions to define the parameters of the circuit:

$$\left.\begin{aligned}\frac{1}{\omega_{cut}R_2C_2}\left(\frac{R_2C_2}{R_1C_1}+\frac{C_2}{C_1}+1-K_0\right) &= b_1 = \sqrt{2}\\ \frac{1}{\omega_{cut}^2R_1R_2C_1C_2} &= b_2 = 1\end{aligned}\right\} \quad (15.62)$$

The time constant of the R_2-C_2 circuit is usually taken at a value close to $1/\omega_{cut}$ [10], [11]. In this case, $\omega_{cut}R_1C_1 \approx 1$ [from the second condition of (15.62)], so that the first condition of (15.62) is reduced to the equation

$$K_0 = 1+\frac{R_2C_2}{R_1C_1}+\frac{C_2}{C_1}-b_1 = 1+1+\frac{C_2}{C_1}-\sqrt{2} \approx 0.59+\frac{C_2}{C_1}$$

Setting $C_2/C_1 = 0.4$ and hence, $R_2/R_1 = 2.5$, we obtain $K_0 \approx 1$. In this example, the operational amplifier is in fact an emitter follower.

For quantitative estimation of the parameters of the low-pass filter, let us set the cutoff frequency f_{cut} at 1000 Hz and the capacitance of the capacitor C_2, at 0.1μF. Then
$C_1 = C_2/0.4 = 0.25\mu F$, $R_1 = 1/\omega_{cut}C_1 \approx 640\ \Omega$
$R_2 = 1/\omega_{cut}C_2 \approx 1600\Omega$

The above example of realization of the second-order filter is just an illustration. For selection of an optimum circuit arrangement and for engineering calculations the reader is invited to study special literature (for example, see [10] and [11]).

15.9. CHEBYSHEV LOW-PASS FILTER

To improve the approximation of the ideal square characteristic of the low-pass filter (Fig. 15.13), the use is often made of the *Chebyshev approximation* in which the square of the Chebyshev polynomial $T_n(x)$ of the respective degree n is employed as the function $F^2(x)$ in formula (15.49). In this case, formula (15.50) is written in the form

$$|K(ix)|^2 = 1/[1+\varepsilon^2 T_n^2(x)] \qquad (15.63)$$

where $x = \omega/\omega_{cut}$.

The coefficient $\varepsilon < 1$ is introduced for limiting the pulsation amplitude of the amplitude-frequency characteristic within the passband, i.e., within the interval $|x| \leqslant 1$. The lower the value of ε, the better the approximation of the amplitude-frequency characteristic in this band but, at the same time, the smaller the slope of the trailing edge of the characteristic curve within the rejection band (for $x > 1$). By varying the value of the coefficient ε and the degree n of the polynomial it is possible to find an acceptable compromise between the contradicting requirements for the approximation of the characteristic within the passband and outside this band.

In Sec. 14.2 it was mentioned that the value of $T_n(x)$ varies within ± 1 within the interval $|x| < 1$ and increases according to $T_n(x) \approx 2^{n-1}x^n$ for $|x| \gg 1$. The graph of the function $|K(ix)|$ for $\varepsilon^2 = 1/5$ and $n = 4$ is plotted in Fig. 15.17.

The pulsation amplitude of the amplitude-frequency characteristic within the passband, i.e.,

$$\Delta K = 1 - 1/\sqrt{1+\varepsilon^2} \qquad (15.64)$$

for low values of ε can be taken approximately equal to $\varepsilon^2/2$ (Fig. 15.17).

Outside the passband, where the values of x are high and $\varepsilon^2 T_n^2(x) \gg \gg 1$, the transfer function decreases monotonously according to

$$|K(ix)| \approx 1/\varepsilon\ |T_n(x)| \qquad (15.65)$$

To compare the Chebyshev approximation of the square amplitude-frequency characteristic with the Butterworth approximation, let

us find the attenuation of the amplitude-frequency characteristic at $x = 3$ for the fourth-order filter, $n = 4$, $\varepsilon^2 = 1/5$. Using the formula given in Sec. 14.2 (or a table of Chebyshev polynomials), we find

$$T_4(3) = 8x^4 - 8x^2 + 1 = 8 \times 3^4 - 8 \times 3^2 + 1 = 574$$

Further,

$$|\mathbf{K}(i3)| \approx \sqrt{5}/574 \approx 4 \times 10^{-3}$$

$$1/|\mathbf{K}(i3)| = 250; \quad (1/|\mathbf{K}(i3)|)_{\text{dB}} = 20 \log_{10} 250 = 20 \times 2.39 \approx 48 \text{ dB}$$

We may see that with the same value of n (4) the attenuation of the amplitude-frequency characteristic of the Chebyshev filter is

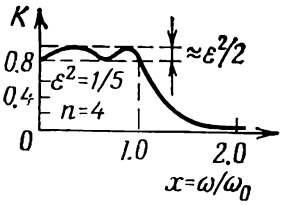


Fig. 15.17. Amplitude-frequency characteristic of a fourth-order Chebyshev filter

8 dB higher than that of the Butterworth filter. Moreover, the Chebyshev filter provides for better approximation of the amplitude-frequency characteristic within the passband: the maximum deviation from unity does not exceed $\varepsilon^2/2 = 0.1$ (over against about 0.3 for the Butterworth filter).

Let us define the poles of the transfer function of the Chebyshev filter (low-pass). As in the preceding section, let us write expression (15.63) in the form

$$|\mathbf{K}(ix)|^2 = \mathbf{K}(p)\,\mathbf{K}(-p)|_{p=ix} = \frac{1}{1+\varepsilon^2 T_n^2(x)} = \frac{1}{B(-p^2)}\Big|_{p=ix} \qquad (15.66)$$

and then find the roots of the equation

$$\varepsilon^2 T_n^2(x) + 1 = 0; \quad T_n(x) = \pm i/\varepsilon \qquad (15.67)$$

Omitting intermediate manipulations (see [6] and [7]), let us write the final expression

$$p_k = \sin \Phi_1 \operatorname{sh} \Phi_2 + i \cos \Phi_1 \operatorname{ch} \Phi_2$$

where Φ_1 and Φ_2 are defined by the formulas

$$\Phi_1 = (2k+1)\frac{\pi}{2n}, \quad k = 0, 1, 2, \ldots, (2n-1) \qquad (15.68)$$

$$\Phi_2 = \frac{1}{n} \operatorname{arcsh}\left(\frac{1}{\varepsilon}\right) \qquad (15.69)$$

For the poles located in the left-hand half-plane p, we obtain the following expression:

$$p_k = -\sin\left[(2k+1)\frac{\pi}{2n}\right]\operatorname{sh}\Phi_2 + i\cos\left[(2k+1)\frac{\pi}{2n}\right]\operatorname{ch}\Phi_2$$

$$k = 0, 1, 2, \ldots, (n-1) \qquad (15.70)$$

Using these poles, let us set up an expression similar to (15.59) for the transfer function $\mathbf{K}(p)$:

$$\mathbf{K}(p) = \frac{b_n}{p^n + b_1 p^{n-1} + b_2 p^{n-2} + \ldots + b_{n-1}p + b_n} = \frac{b_n}{(p-p_1)(p-p_2)\ldots(p-p_n)} \qquad (15.71)$$

In contrast to the Butterworth filter, the coefficient b_n is not equal to unity (because the poles of the transfer function are located on an ellipse and not on the unit-radius circle). Therefore, the term b_n is introduced into the numerator for normalizing the amplitude-frequency characteristic to unity at $\omega = 0$ (and, respectively, $p = 0$).

The numerical values of the coefficients $b_1, b_2, \ldots, b_n$, and also of the poles $p_1, p_2, \ldots, p_n$, as a function of the degree n and the nonuniformity coefficient ε of the amplitude-frequency characteristic are given in the literature on filter calculation [7], [8], [12].

To illustrate the synthesis of a Chebyshev filter, let us determine the circuit arrangement and parameters of the filter, given the following conditions: the nonuniformity within the passband is not higher than 3 dB, and the attenuation for $x = \omega/\omega_{cut} = 4$ not less than 30 dB. For the given nonuniformity, setting $\Delta K = 1 - 1/\sqrt{2}$ in expression (15.64), we obtain $\varepsilon = 1$. Further, using formula (15.65), we find the required value

$$|T_n(x)| = |T_n(4)| \geqslant 1/|\mathbf{K}(i4)|$$

A 32-dB attenuation corresponds to a decrease of the amplitude-frequency characteristic by a factor $\sqrt{1000} \approx 32$. With the maximum value of the amplitude-frequency characteristic equal to unity, we obtain the following conditions for determining the order of the Chebyshev polynomial: $T_n(4) \geqslant 1/32$. By looking over the first three lower-degree polynomials (see Sec. 14.2), we make certain that the second-degree polynomial, which for $x = 4$ is equal to $T_2(4) = 2x^2 - 1 = 31$, ensures the required rate of decrease of the amplitude-frequency characteristic in the rejection band. Using formulas (15.69) and (15.70), we find

$$\Phi_2 = \frac{1}{n}\operatorname{arcsh}\left(\frac{1}{\varepsilon}\right) = \frac{1}{2}\operatorname{arcsh} 1 = 0.44$$

$$p_1 = -\sin\left(\frac{\pi}{2\times 2}\right)\operatorname{sh} 0.44 + i\cos\left(\frac{\pi}{2\times 2}\right)\operatorname{ch} 0.44 = -0.322 + i0.777$$

$$p_2 = p_1^* = -0.322 - i0.777$$

The transfer function (by formula 15.71) is

$$K(p) = \frac{b_2}{(p-p_1)(p-p_1^*)} = \frac{b_2}{p^2+b_1p+b_2} = \frac{b_2}{p^2+0.645p+0.708} \qquad (15.71')$$

Setting (as in the preceding section) the coefficients of the polynomial in the denominator of expression (15.61) $b_1 = 0.645$ and $b_2 = 0.708$, respectively, we obtain the following expressions for determining the parameters of the active R-C circuit:

$$\frac{1}{\omega_{cut}R_2C_2}\left(\frac{R_2C_2}{R_1C_1} + \frac{C_2}{C_1} + 1 - K_0\right) = b_1 = 0.645$$

$$\frac{1}{\omega_{cut}^2 R_1R_2C_1C_2} = b_2 = 0.708$$

Preserving the relations adopted in Sec. 15.8 for the Butterworth filter ($\omega_{cut}R_2C_2 \approx 1$, $C_2/C_1 = 0.4$), we obtain

$$\omega_{cut}R_1C_1 = \frac{1}{\omega_{cut}R_2C_2}\frac{1}{b_2} = \frac{1}{b_2} = \sqrt{2}$$

$$K_0 = 1 + \frac{C_2}{C_1} + \frac{1}{\sqrt{2}} - 0.645 \approx 1.46$$

From comparison between these results and those obtained for the Butterworth filter, it is seen that, by varying the amplification factor K_0 of the operational amplifier and insignificantly changing the resistances of the resistors R_1 and R_2 (or the capacitances of the capacitors C_1 and C_2) one can change over from the Butterworth filter to the Chebyshev filter. However, it should be noted that for $n = 2$ and $x = \omega/\omega_{cut} = 4$ the Butterworth filter ensures only a 24-dB attenuation of the amplitude-frequency characteristic [see formula (15.52) for $n = 2$ and $y = 2$]. A 30-dB attenuation would require $n \geqslant 3$ (one second-order link and one aperiodic link). This advantage of the Chebyshev filter in the rejection band (faster decrease of the amplitude-frequency characteristic) is attained at the cost of some deterioration of uniformity within the passband of the filter.

15.10. SYNTHESIS OF VARIOUS FILTERS ON THE BASIS OF A LOW-PASS FILTER

Let us return to function (15.50) approximating the square amplitude-frequency characteristic of an ideal low-pass filter and introduce a new variable

$$\nu = 1/x = \omega_{cut}/\omega \qquad (15.72)$$

Then

$$K_{lp}^2(x) = 1/[1 + (1/\nu)^{2n}] = \nu^{2n}/(1+\nu^{2n}) = K_{hp}^2(\nu) \qquad (15.73)$$

The new function $K_{hp}(\nu)$, obtained from the amplitude-frequency characteristic $K_{lp}(x)$ of the low-pass filter by substituting $\nu = 1/x$ for the argument, is shown in Fig. 15.18 (for $n = 2$). The function K_{hp} can be regarded as the amplitude-frequency characteristic of a hing-pass Butterworth filter, having the same nonuniformity in the passband $1 < \nu < \infty$ as the function $K_{lp}(x)$ in the band $0 <$

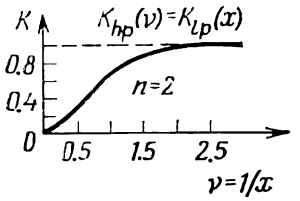


Fig. 15.18. High-pass Butterworth filter

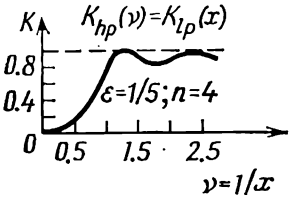


Fig. 15.19. High-pass Chebyshev filter

$< x < 1$. Thus, when synthesizing a high-pass filter one can use the approximating function $F(x)$ [see (15.49)] after replacing its argument by $\nu = 1/x$.

In accordance with this replacement, the frequency variable p in (15.58) should be replaced by $s = 1/p$. Function (15.58) in this case assumes the form

$$K_{hp}(s) = \frac{1}{(1/s)^2 + \sqrt{2}(1/s) + 1} = \frac{s^2}{s^2 + \sqrt{2}s + 1} \qquad (15.74)$$

The poles of the transfer function, $K_{hp}(s)$, i.e., the roots of the equation $s^2 + \sqrt{2}s + 1 = 0$, remain the same as in (15.58).

In a similar way, one can obtain the transfer function of the high-pass Chebyshev filter.

The relation between the amplitude-frequency characteristics of the low- and high-pass Chebyshev filters is illustrated in Fig. 15.19 (for $n = 4$).

Other filters, such as band-pass filters can be synthesized by respectively transforming the variable p in the transfer function of the original low-pass filter [7], [8], [12].

15.11. SENSITIVITY OF CIRCUIT CHARACTERISTICS TO CHANGES IN THE PARAMETERS OF THE CIRCUIT ELEMENTS

In Ch. 5 we briefly discussed the effect of a change in the amplification factor K_a and in the modulus K_{fb} of the transfer function of the feedback circuit on the gain of the whole system. The question of the margin of stability of a two-port network with a feedback loop was not discussed. This question has become of special importance in connection with the synthesis of active R-C circuits (and integra-

ted circuits), in which one has to reckon with the comparatively high instability of the active element parameters and their dependence on the operating conditions (supply voltage fluctuations, temperature variations, etc.). The scattering of the parameters of resistors, capacitors and, particularly, of active circuit components (transistors, microelectronic devices, etc.) should also be taken into consideration.

The transfer function of a linear circuit is unambiguously defined by its poles and zeros on the p-plane. Therefore, it is only logical to evaluate the various destabilizing factors in terms of the shift of the poles and zeros caused by them.

Special attention has to be paid to the shift of the poles, since it is their position that defines such important characteristics as the margin of stability, gain, and resonance frequency.

The mathematical definition of the sensitivity of a circuit parameter W to a change of an element parameter X, given in the form of the equation

$$S_X^W = \frac{dW/W}{dX/X} = \frac{X}{W}\frac{dW}{dX} \tag{15.75}$$

has found wide application [13].

The symbol W may have the meaning of the amplification factor K_a, Q-factor, root $p_k = \sigma_k + i\omega_k$, etc., while X may imply the resistance of a resistor, capacitance of a capacitor, or any parameter of an active element (slope of a characteristic curve, internal conductance, etc.).

As applied to the evaluation of the *pole sensitivity*, expression (15.75) is often written in the form

$$S_X^{p_k} = \frac{d\sigma_k/\sigma_k}{dX/X} + i\frac{d\omega_k/\omega_k}{dX/X} = \frac{X}{\sigma_k}\frac{d\sigma_k}{dX} + i\frac{X}{\omega_k}\frac{d\omega_k}{dX} \tag{15.76}$$

The first term on the right-hand side characterizes the pole shift parallel to the real axis of the plane $p = \sigma + i\omega$, while the second, that parallel to the axis $i\omega$. Thus, the first term can serve as an estimate of the margin of the circuit stability (since the approach of $|\sigma_k|$ to zero means the loss of stability), the second term characterizing the change in the frequency corresponding to the peak of the amplitude-frequency characteristic of the circuit.

Let us illustrate the use of formula (15.76) with an example of the transfer function $\mathbf{K}(p)$ of the Butterworth filter considered in Sec. 15.8. Let us define the pole sensitivity of function (15.61) to a change in K_0 regarded as the parameter X in (15.76).

Substituting into (15.61) $\omega_{cut}R_1C_2 = \omega_{cut}R_2C_1 = 1$, $C_1R_2/C_2R_1 = 1$, and $C_1/C_2 = 0.4$ (see the end of Sec. 15.8), we obtain

$$\mathbf{K}(p) = \frac{K_0}{p^2 + (2.4 - K_0)\,p + 1} \tag{15.77}$$

The poles of the function **K** (p) are

$$p_{1,2} = -\frac{2.4-K_0}{2} \pm \sqrt{\left(\frac{2.4-K_0}{2}\right)^2 - 1}$$

For $K_0 < 0.4$, the roots are real, while for $K_0 > 0.4$, they are complex.

In the last case

$$p_{1,2} = -\frac{2.4-K_0}{2} \pm i\sqrt{1-\left(\frac{2.4-K_0}{2}\right)^2}$$

As in (15.61), we imply the normalized variable $p = (\sigma + i\omega)/\omega_{cut}$.

The modulus of $p_{1,2}$ is equal to unity (for $K_0 > 0.4$), while the real and imaginary parts are respectively

$$\left.\begin{aligned} \frac{\sigma_{1,2}}{\omega_{cut}} &= \mathrm{Re}\,(p_{1,2}) = -\frac{2.4-K_0}{2} \\ \frac{\omega_{1,2}}{\omega_{cut}} &= \mathrm{Im}\,(p_{1,2}) = \pm\sqrt{1-\left(\frac{2.4-K_0}{2}\right)^2} \end{aligned}\right\} \qquad (15.78)$$

As K_0 increases, starting from $K_0 = 0.4$, the poles $p_{1,2}$ move along the unit-radius circle (Fig. 15.20), p_1 moving clockwise and p_2, counterclockwise. At $K_0 = 2.4$ the coefficient multiplying p in the denominator of (15.77) vanishes, this meaning the loss of the circuit stability.

Differentiating expressions (15.78) and substituting the derivatives

$$\frac{d\left(\frac{\sigma_1}{\omega_{cut}}\right)}{dK_0} = \frac{1}{2}, \quad \frac{d\left(\frac{\omega_1}{\omega_{cut}}\right)}{dK_0}$$

$$= \frac{2.4-K_0}{4\sqrt{1-\left(\frac{2.4-K_0}{2}\right)^2}}$$

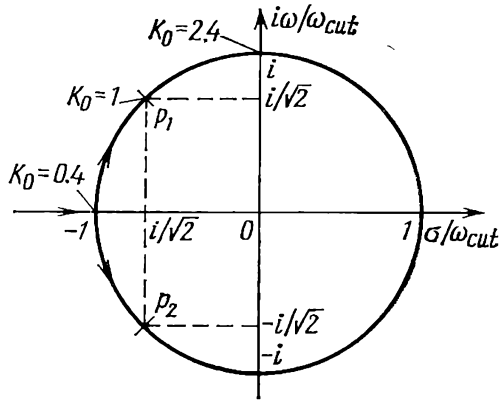


Fig. 15.20. Movement of poles along the unit-radius circle with a change of K_0

and also σ_1/ω_{cut} and ω_1/ω_{cut} into formula (15.76) for $k = 1$, i.e., for the root p_1, we obtain

$$S_{K_0}^{p_1} = \frac{K_0}{K_0-2.4} + i\,\frac{K_0(2.4-K_0)}{4-(2.4-K_0)^2}$$

For the initial (design) value of K_0 equal to unity

$$S_{K_0}^{p_1} = -\frac{1}{1.41} + i\,\frac{1.4}{4-1.41^2} = -0.707 + i0.707$$

However, for higher values of K_0, e.g., for $K_0 = 2$,

$$S_{K_0}^{p_1} = \frac{2}{2-2.4} + i\frac{2(2.4-2)}{4-(2.4-2)^2} \approx -5 + i0.2$$

This means that the sensitivity of the pole abscissa reaches —5, while the sensitivity of the ordinate decreases to 0.2.

As K_0 approaches the critical value corresponding to the start of generation, the sensitivity of the pole abscissa sharply increases, while that of the ordinate drops down.

The required shape of the amplitude-frequency characteristic with simultaneous increase of the gain in the filter can be achieved by redistributing the capacitances C_1 and C_2 (reducing C_1 and increasing C_2), which requires an increase in K_0 [since in (15.61) the coefficient K_0 is in the numerator]. However, in this case the pole sensitivity of the transfer function to changes in K_0 increases.

15.12. SIMULATION OF INDUCTANCE BY MEANS OF AN ACTIVE R-C CIRCUIT. GYRATOR

The traditional theory and practice of frequency-selective electric circuit deals with three elements: inductance coils, capacitors and resistors. The development of microelectronics, based on integrated circuits, has resulted in the appearance of film capacitors

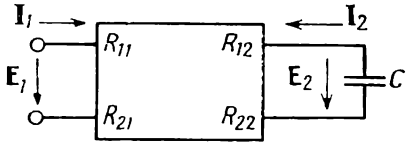


Fig. 15.21. Simulation of inductance by means of an active two-port network loaded by a capacitor

and resistors, while inductance coils for integrated circuits have not yet been realized (at least in a sufficiently wide range of L with high requirements for the Q-factor of the coils). Ordinary inductance coils are technologically incompatible with integrated circuits. In this connection, in modern radioelectronics much attention is being paid to the methods of constructing frequency-selective networks on the basis of resistors and capacitors.

One of possible methods of solving this problem is simulation of an inductance by means of an active two-port network composed of resistive elements and loaded by a capacitor (Fig. 15.21).

The problem may be worded as follows: what must be the parameters R_{11}, R_{12}, R_{21}, and R_{22} to make the input impedance of the two-port network have the character of an inductive reactance $i\omega L$, with the load impedance being $Z_l(i\omega) = 1/i\omega C$?

A reply to this question can be obtained from expression (5.23) by equating $Z_{in}(i\omega)$ in it to the required quantity $i\omega L$ and equating Z_l to $1/i\omega C$:

$$Z_{in}(i\omega) = R_{11} - \frac{R_{12}R_{21}}{R_{22} + 1/i\omega C} = i\omega L$$

If the conditions $R_{11} = R_{22} = 0$ and $R_{12}R_{21} = -L/C$ are satisfied, we obtain $Z_{in} = i\omega L$ and

$$L = -R_{12}R_{21}C \tag{15.79}$$

Since L and C are positive, one of the factors R_{12} or R_{21} must be positive and the other, negative.

Denoting $-R_{12} = R_1$ and $R_{21} = R_2$, we come to the following R-matrix of the inductance-simulating two-port network:

$$[R] = \begin{bmatrix} 0 & -R_1 \\ R_2 & 0 \end{bmatrix} \tag{15.80}$$

Equations (5.4) corresponding to this matrix assume the form

$$\mathbf{E}_1 = -R_1\mathbf{I}_2, \quad \mathbf{E}_2 = R_2\mathbf{I}_1 \tag{15.81}$$

The two-port network described by matrix (15.80) is referred to as the *gyrator* [10], [11].

When $R_1 = R_2$, the gyrator does not consume power, nor does it produce power. This follows directly from the energy balance of the circuit shown in Fig. 15.21 (with $R_{11} = R_{22} = 0$, $R_{12} = -R$, and $R_{21} = R$):

$$\mathbf{E}_1\mathbf{I}_1 = (-R_1\mathbf{I}_2)\,\mathbf{I}_1, \quad \mathbf{E}_2\mathbf{I}_2 = (R_2\mathbf{I}_1)\,\mathbf{I}_2$$
$$\mathbf{E}_1\mathbf{I}_1 + \mathbf{E}_2\mathbf{I}_2 = \mathbf{I}_1\mathbf{I}_2\,(R_2 - R_1)$$

When $R_1 = R_2$, $\mathbf{E}_1\mathbf{I}_1 + \mathbf{E}_2\mathbf{I}_2 = 0$. Thus, the gyrator, though nonreciprocal, is a passive two-port network. (When $R_1 \neq R_2$, the gyrator is sometimes referred to as active.)

Thus, the R-matrix and the equations of the ideal passive gyrator are as follows:

$$\left.\begin{aligned} [R] &= \begin{bmatrix} 0 & -R \\ R & 0 \end{bmatrix} \\ \mathbf{E}_1 &= -R\mathbf{I}_2, \quad \mathbf{E}_2 = R\mathbf{I}_1 \end{aligned}\right\} \tag{15.82}$$

Correspondingly, the Y-matrix and the gyrator equations in terms of admittances are

$$\left.\begin{aligned} [Y] &= \begin{bmatrix} 0 & -Y \\ Y & 0 \end{bmatrix} \\ \mathbf{I}_1 &= -Y\mathbf{E}_2, \quad \mathbf{I}_2 = Y\mathbf{E}_1 \end{aligned}\right\} \tag{15.83}$$

where $Y = 1/R$.

From expressions (15.82) and (15.83) it follows that *two dependent sources* are necessary to realize a gyrator.

In the first case, where the gyrator is constructed on the basis of its R-matrix, two dependent voltage sources ($R\mathbf{I}_1$ and $-R\mathbf{I}_2$) are required, which are controlled by the currents $\mathbf{I}_1$ and $\mathbf{I}_2$, respectively.

In the second case, where the Y-matrix is used, two dependent current sources ($Y\mathbf{E}_1$ and $-Y\mathbf{E}_2$) are required, which are controlled by the voltages $\mathbf{E}_1$ and $\mathbf{E}_2$, respectively.

Use can also be made of circuits based on the utilization of dependent voltage-controlled voltage sources or current-controlled current sources.

Various electron devices can be used as nonreciprocal (active) elements. Shown in Fig. 15.22*a* is the equivalent circuit of an ideal

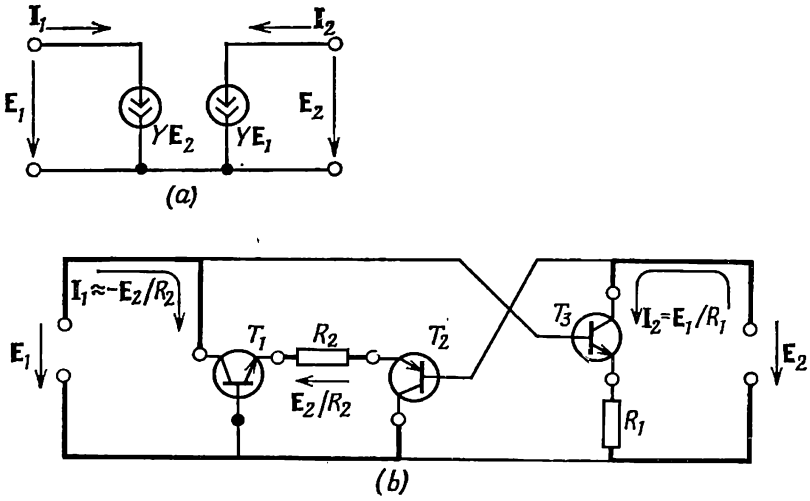


Fig. 15.22. Gyrator built around two voltage-controlled current sources
(*a*) equivalent circuit; (*b*) circuit diagram

gyrator based on two voltage-controlled current sources. This circuit is obtained from the one of Fig. 5.2*a* at $Y_{11} = Y_{22} = 0$.

One of possible versions of a transistorized gyrator is shown in Fig. 15.22*b*. (The d-c sources are not shown.)

The voltage $\mathbf{E}_1$ is applied simultaneously across the collector-to-base junction of the transistor T_1 and across the base-to-emitter junction of the transistor T_3. In the first transistor, the current due to the voltage $\mathbf{E}_1$ is negligible (because of the high resistance of the collector-to-base circuit), while in the second transistor, the voltage $\mathbf{E}_1$ is actually applied across the resistor R_1 (because of the low resistance of the emitter-to-base circuit), thus giving rise to a current $\mathbf{E}_1/R_1$ in the emitter circuit. With $\alpha \to 1$, this current coincides with the collector current. Thus, the transistor T_3, together with the resistor R_1, behaves like a dependent current source $\mathbf{I}_2 = \mathbf{E}_1/R_1$.

On the other hand, the voltage $\mathbf{E}_2$ is applied across the resistor R_2 through the base-to-emitter junctions of the transistors T_2 and T_1 connected in series.

The current through R_2, approximately equal to $\mathbf{E}_2/R_2$ and directed from right to left (with the direction of $\mathbf{E}_2$ being as shown in Fig. 15.22*b*), is equal (with $\alpha \to 1$) to the collector currents of the transistors T_1 and T_2. As a result, we have a closed circuit of a dependent current source $\mathbf{I}_1 = -\mathbf{E}_2/R_2$ shown in Fig. 15.22*b* by a

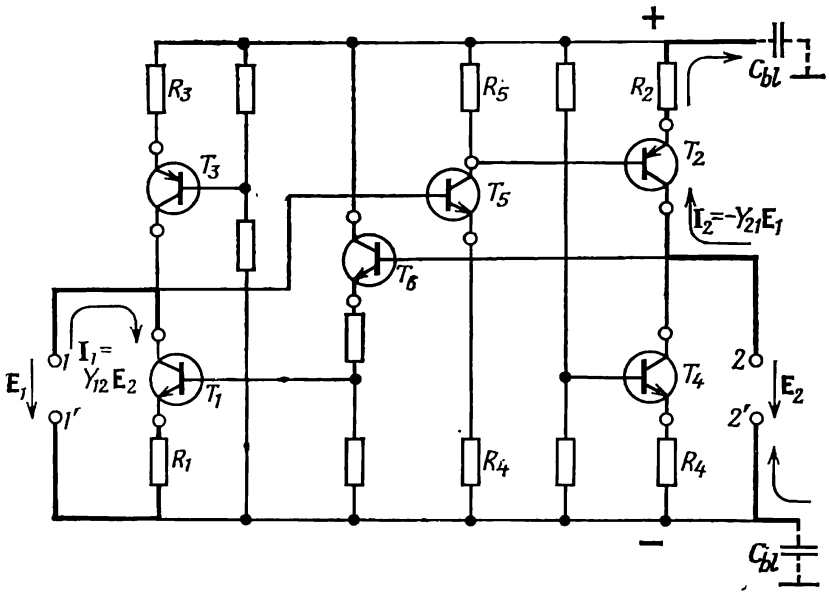


Fig. 15.23. Improved gyrator circuit

thick line. The minus sign takes into account the direction of the current adopted in the circuit in Fig. 15.22*a* and the reverse direction of the current through the resistor R_2.

When the shunting effect of the power supply circuits is taken into account, additional admittances appear, which reduce the input and output impedances of the dependent current sources and result in that the elements Y_{11} and Y_{22} in the Y-matrix of the gyrator do not vanish. Therefore, the simple circuit shown in Fig. 15.22*b* has not gained wide application.

The circuit shown in Fig. 15.23 is essentially free from the above disadvantages. The collector circuit of the transistor T_1 is fed through the transistor T_3 connected in series. The sufficiently high differential resistance of the collector-to-emitter junction of this transistor grows still higher due to the negative feedback provided by the resistor R_3 in the emitter circuit. With $R_3 = R_1$, the voltage of the power source is divided equally between the transistors T_1 and T_3.

A similar method is used in the power-supply circuit of the transistor T_2 (by means of the transistor T_4 and the resistor R_4).

The voltage $\mathbf{E}_1$ is applied through a common-emitter amplifier (transistor T_5) across the base and emitter terminals of the transistor T_2. Since a common-emitter amplifier shifts the oscillation phase by 180°, the current in the collector circuit of the transistor T_2 is $\mathbf{I}_2 = -Y_{21}\mathbf{E}_1$ (the circuit is closed through the blocking capacitor shunting the power source).

On the other hand, the voltage $\mathbf{E}_2$ applied to the transistor T_1 through the emitter follower T_6 (without phase shift) produces a current $\mathbf{I}_1 = Y_{12}\mathbf{E}_2$ in the collector circuit of the transistor T_1. When a capacitor C is connected to the terminals 2—2′ to obtain the required inductance L (between the terminals 1—1′), by $\mathbf{E}_2$ one should imply the voltage produced across the capacitor C by the current $\mathbf{I}_2$.

Thus, the required gyrator matrix is realized. Various gyrator circuits are known in the art [10], [11].

15.13. SOME SPECIFIC FEATURES OF THE SYNTHESIS OF DIGITAL FILTERS

The synthesis of both digital and analog filters is usually based on their transfer function represented in the form of a rational fraction (13.64). After the corresponding approximation of the given transfer function $\hat{\mathbf{K}}(z)$, one determines the position of the zeros and poles on the z-plane and then finds the weighting coefficients a_n and b_m entering into the polynomials in expression (13.63).

A digital filter can be realized either as a combination of simple links (of the first or second order) or in the form of the canonical circuit described in Sec. 13.7 (Fig. 13.14).

When a filter is subdivided into simple links, all the limitations noted in Sec. 15.4 for analog circuits fall away. As for digital circuits, the question of matching the input, output and load impedances, as well as the question of decoupling individual links, does not arise at all. That is why the parallel connection of simple links is widely used together with their cascade (series) connection.

In the second case, function (13.64) is written in the form of a product of simple factors each of which is the transfer function of a link (see similar subdivision in Sec. 15.4). In the first case, function (13.64) is decomposed into partial fractions:

$$\hat{\mathbf{K}}(z) = A_0 \frac{P(z)}{Q(z)} = A_0 \sum_{k=1}^{m} \frac{\mathbf{A}_k}{z - z_{pk}}$$

where $\mathbf{A}_k = \frac{P(z)}{[dQ(z)/dz]_{z=z_{pk}}}$ is the residue of the function $\hat{\mathbf{K}}(z)/A_0$ at the pole z_{pk}.

If the denominator $Q(z)$ has a total of m roots, of which there are m_1 real roots (those lying on the real axis) and m_2 complex-conjugate pairs of roots ($m = m_1 + 2m_2$), then

$$\hat{\mathbf{K}}(z) = A_0\left[\sum_{k=1}^{m_1}\frac{\mathbf{A}_k}{z-z_{pk}} + \sum_{k=1}^{m_2}\left(\frac{\mathbf{A}_k}{z-z_{pk}} + \frac{\mathbf{A}_k^*}{z-z_{pk}^*}\right)\right]$$

This expression is easily reduced to the form

$$\hat{\mathbf{K}}(z) = A_0\left[\sum_{k=1}^{m_1}\frac{\mathbf{A}_k}{z-z_{pk}} + \sum_{k=1}^{m_2}\frac{a_{0k}z+a_{1k}}{z^2-b_{1k}z-b_{2k}}\right]$$

where $a_{0k} = 2\mathrm{Re}(\mathbf{A}_k)$, $a_{1k} = -2\ \mathrm{Re}(z_{pk})\ \mathrm{Re}(\mathbf{A}_k)$, $b_{1k} = 2\mathrm{Re}(z_{pk})$, and $b_{2k} = |z_{pk}|^2$.

In both the cascade and parallel circuits, individual links are realized according to the circuit diagram described in Sec. 13.11-3 (Fig. 13.24). The weighting coefficients of the second-order link are defined by formula (13.76), while those of the first-order link are found directly from its transfer function.

The approximation of a specified characteristic of a digital filter is often carried out by the method based on the use of the results of approximation of the corresponding analog filter. The essence of this method consists in the following. Let us have a physically realizable analog filter whose transfer function $\mathbf{K}(p)$ satisfies the requirements for this filter and let it be necessary to find an equivalent (to a certain extent) discrete filter. The poles and zeros of the function $\mathbf{K}(p)$ are supposed to be known.

Since the sought for transfer function $\hat{\mathbf{K}}(z)$ of the digital filter is unambiguously defined by its poles and zeros on the z-plane and the relation between z and p is $z = e^{pT}$ (see Sec. 13.8), at first glance it would seem logical to use expression $z_{0n} = e^{p_{0n}T}$, $z_{pm} = e^{p_{pm}T}$. But actually, such an approach can under certain conditions be incorrect. In Sec. 13.11-2 it was shown by an example of an R-C circuit that the use of the transform $z = e^{pT}$ yields a discrete impulse response $g_T(t)$ coinciding with the samples of the continuous characteristic $g(t)$ of the analog circuit; however, the amplitude-frequency characteristic is in this case completely distorted. A similar result is also obtained for more involved circuits. In this connection, the synthesis method based on the transform $z = e^{pT}$ is known as the *invariant method* with respect to the *impulse response* of the filter.

To effect a synthesis that would be *invariant with respect to the amplitude-frequency* characteristic, one should use a transformation

whereby the entire imaginary axis $i\omega$ of the p-plane would be mapped onto the z-plane in a *single circuit* round the circle of radius $|z| = 1$.

This requirement is satisfied by the *bilinear* (fractionally rational) transformation

$$z = (1 + p)/(1 - p), \quad p = (z - 1)/(z + 1) \tag{15.84}$$

where $p = (\sigma + i\omega)/\Omega_0$, and Ω_0 is an arbitrary constant that makes the quantity p dimensionless and is selected for normalization reasons.

To reveal the essence of the bilinear transformation, let us set $\sigma = 0$, i.e., $p = i\omega/\Omega_0$, and, according to (15.84), write

$$z|_{p=i\omega/\Omega_0} = \frac{1+i\omega/\Omega_0}{1-i\omega/\Omega_0} = e^{-i2\arctan \omega/\Omega_0} = e^{i\varphi(\omega/\Omega_0)} \tag{15.85}$$

From this expression it follows that the movement of the point p along the axis $i\omega/\Omega_0$ corresponds to the movement of the point z along the circuit of radius $|z| = 1$. In this respect the bilinear transformation does not differ from the ordinary z-transformation in which $z|_{p=i\omega} = e^{i\omega T}$ (see Sec. 13.8). The only difference consists in that, unlike the angle ωT which increases in proportion to the frequency ω, the angle $\varphi(\omega/\Omega_0) = 2 \arctan \omega/\Omega_0$ in the bilinear transformation rises *nonlinearly*: as ω tends to $\pm\infty$, the angle $\varphi(\omega/\Omega_0)$ tends to its limiting values of $\pm\pi$. Thus, the whole axis $i\omega/\Omega_0$ of the p-plane is mapped onto the z-plane in a single circuit round the circle of radius $|z| = 1$ and a reciprocal one-to-one correspondence between the points p and z is ensured for the entire p-plane.

Comparison between the functions $e^{i\varphi(\omega/\Omega_0)}$ and $e^{i\omega T}$ enables one to treat $\varphi(\omega/\Omega_0) = 2 \arctan \omega/\Omega_0$ as an equivalent frequency $\omega_d T$ (dimensionless) which is related to the ordinary frequency ω used in the analysis and synthesis of analog circuits by the relation

$$\omega_d T = 2 \arctan \omega/\Omega_0 \tag{15.86}$$

Correspondingly,

$$\frac{\omega}{\Omega_0} = \tan\left(\frac{\omega_d T}{2}\right) \tag{15.87}$$

The normalizing frequency Ω_0 can be determined by establishing a relation between some characteristic frequencies of the transfer functions of the analog and digital circuits. For example, if we have to synthesize a digital filter with a given cutoff frequency $\omega_{d,cut}$, that is equivalent (as to the amplitude-frequency characteristic) to an analog filter with a cutoff frequency ω_{cut}, expressions (15.86)

and (15.87) may be written as

$$\left.\begin{aligned} \omega_{d,\,cut}T &= 2 \arctan \frac{\omega_{cut}}{\Omega_0} \\ \frac{\omega_{cut}}{\Omega_0} &= \tan\left(\frac{\omega_{d,\,cut}}{2}\right) \end{aligned}\right\} \qquad (15.86')$$

From the last expression it follows that

$$\Omega_0 = \frac{\omega_{cut}}{\tan\left(\frac{\omega_{d,\,cut}T}{2}\right)} \qquad (15.87')$$

Let, for example, it be necessary that the cutoff frequency of the digital filter should be 10% of the sampling frequency $1/T$. Then $\omega_{d,\,cut}T = 0.1 \times 2\pi$ and

$$\tan\left(\frac{\omega_{d,\,cut}T}{2}\right) = \tan\left(\frac{0.1\times 2\pi}{2}\right) = \tan 18^\circ = 0.3249$$

so that expression (15.86) is reduced to

$$\omega_d T = 2 \arctan (0.3249\omega/\omega_{cut}) = 2 \arctan (0.3249x) \qquad (15.88)$$

where $x = \omega/\omega_{cut}$ is the normalized frequency used when approximating the amplitude-frequency characteristic of the analog filter (see Secs. 15.8 and 15.9).

The relation between $\omega_d T$ and x thus obtained makes it possible to construct the amplitude-frequency characteristic of the digital filter being synthesized, using the given characteristic of the original analog circuit. As an example of such a specified characteristic, Fig. 15.24*a* shows the amplitude-frequency characteristic of the Chebyshev filter (with $n = 2$), calculated in Sec. 15.9 by formula

$$|\mathbf{K}(ix)| = 1/\sqrt{1 + T_2^2(x)} \qquad (15.71')$$

The amplitude-frequency characteristic of the digital filter is shown in Fig. 15.24*b*. It is seen that this characteristic, while retaining its scale along the ordinate axis, is compressed along the abscissa axis within the limits $-\pi \leqslant \omega_d T < \pi$.

Now, let us find the structure and parameters of the digital filter being synthesized.

Proceeding from the transfer function of the original analog filter [see (15.71')]

$$\mathbf{K}(p_x) = \frac{c}{(p_x - p_{xp1})(p_x - p_{xp1}^*)} \qquad (15.89)$$

where $p_x = \frac{\sigma + i\omega}{\omega_{cut}}$ is the variable normalized with respect to the cutoff frequency, and transforming to a new variable $p = \frac{\sigma + i\omega}{\Omega_0} =$

$= \tan\left(\frac{\omega_{d,cut}T}{2}\right) p_x = 0.3249p_x$, we obtain on the basis of (15.84)

$$p = 0.3249p_x = \frac{z-1}{z+1}, \qquad p_x = \frac{1}{0.3249}\,\frac{z-1}{z+1}$$

Let us substitute the above expression for p_x into (15.89). The poles p_{xp1} and p^*_{xp1}, as in the example given in Sec. 15.9, are equal to $-0.322 + i0.777$ and $-0.322 - i0.777$, respectively.

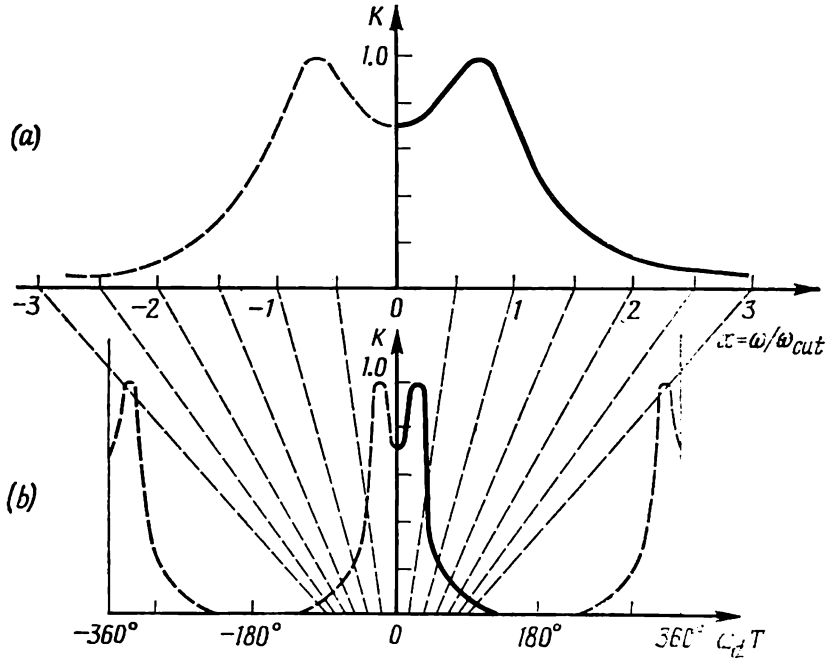


Fig. 15.24. Amplitude-frequency characteristics of (*a*) an analog filter and (*b*) the respective digital filter

After simple manipulations, we obtain the following result:

$$\hat{\mathbf{K}}(z) = \frac{A_0(z+1)^2}{(z-z_{p1})(z-z^*_{p1})} = \frac{A_0(1+2z^{-1}+z^{-2})}{1-b_1z^{-1}-b_2z^{-2}}$$

where $b_1 = 2\,\mathrm{Re}\,(z_{p1})$ and $b_2 = -\,|\,z_{p1}\,|^2$.

The poles of the function $\hat{\mathbf{K}}(z)$ on the z-plane are related to the poles $p_{xp1,2}$ by the relations

$$z_{p1} = \frac{1+0.3249p_{xp1}}{1-0.3249p_{xp1}} = 0.72 + i0.393$$

$$z_{p2} = z^*_{p1} = 0.72 - i0.393$$

Thus, the use of the bilinear z-transformation has resulted in the appearance of a double zero in the transfer function (at the point $z = -1$).

The structure of the filter coincides with that shown in Fig. 13.24. In this case, the feedback coefficients (see Sec. 13.11-3) are

$$b_1 = 2\mathrm{Re}\,(z_{p1}) = 2 \times 0.72 = 1.44$$

$$b_2 = -|z_{p1}|^2 = -(0.821)^2 = -0.674$$

and the feedforward coefficients, $a_0 = 1$, $a_1 = 2$ and $a_2 = 1$.

The constant A_0 is introduced for normalization purposes. With $\omega_d = 0$, $z = 1$ and the function $|\hat{\mathbf{K}}(z)|$, by hypothesis, must be equal to unity, and so must be the function $\mathbf{K}(p)$ when $\omega = 0$. With the coefficients a_i and b_i being as indicated above, $A_0 = 0.0585$.

When synthesizing a digital filter, of great importance is the choice of the number of digits in the analog-to-digital converter and in the arithmetic unit, based on the admissible level of the quantization and rounding-off noise (see Secs. 13.12 and 13.14).

It is different with the weighting coefficients b_1 and b_2, because accurate representation of these coefficients in the binary system may require a considerable number of digits (1.011101 for b_1 and 0.10101101 for b_2). However, the number of digits can usually be considerably reduced at the cost of an insignificant deviation of the amplitude-frequency characteristic from the specified one. For example, by reducing the number of digital positions for b_1 to 1.0111 (1.4375) and for b_2 to 0.1011 (0.687) one can obtain an amplitude-frequency characteristic practically coinciding with the specified one.

APPENDIX I

SIGNAL WITH MINIMUM PRODUCT OF DURATION AND FREQUENCY BAND

In Sec. 2.10 it was noted that the product of the duration of a signal and its frequency band cannot be smaller than some definite value. Let us find this value and the optimum signal having the minimum possible product of the duration and frequency band.

First, we should define the terms signal duration and spectrum width. Used in practice are different definitions whose selection depends on the application of the signal and on its shape and spectrum. In some cases, the choice is arbitrary. For example, the duration of a signal having a rectangular shape can naturally be defined by the rectangle base, however the spectrum width is defined either as the base of the main lobe (as, for example, in Sec. 2.9-1) or at the level of $1/\sqrt{2}$ of the maximum value. The duration of the Gaussian pulse (Sec. 2.9-2) and the width of its spectrum are defined at the level of 0.606 of the maximum value. Sometimes, an energy criterion is used, in which case by the spectrum width is understood the frequency band containing a specified portion of the total energy of the signal.

To define the limiting relations between the duration and the frequency band, the so-called *method of moments* has found wide application in the modern theory of signals.

By analogy with the moment of inertia in mechanics, the effective duration of a signal $s(t)$ can be defined by the expression

$$T_e^2 = \int_{-\infty}^{\infty} (t-t_0)^2 s^2(t)\, dt \Big/ \int_{-\infty}^{\infty} s^2(t)\, dt \tag{I.1}$$

where the pulse centre t_0 is found from the condition

$$t_0 = \int_{-\infty}^{\infty} t s^2(t)\, dt \Big/ \int_{-\infty}^{\infty} s^2(t)\, dt$$

We imply that the function $s(t)$ is quadratically integrable (signal of finite energy).

In a similar way, the effective spectrum width $\Omega_e = 2\pi F_e$ if defined by the expression*

$$\Omega_e^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 S^2(\omega)\, d\omega \Big/ \frac{1}{2\pi} \int_{-\infty}^{\infty} S^2(\omega)\, d\omega \tag{I.2}$$

Since the modulus of the spectrum $S(\omega)$ is independent of the time shift of $s(t)$, we may set $t_0 = 0$. Finally, the signal $s(t)$ can be normalized so that

* Here we imply signals without radio-frequency carrier.

its energy E is equal to unity and consequently,

$$\int_{-\infty}^{\infty} s^2(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} S^2(\omega)\,d\omega = 1$$

Subject to these conditions, the expressions for T_e and Ω_e assume the form

$$T_e^2 = \int_{-\infty}^{\infty} t^2 s^2(t)\,dt, \qquad \Omega_e^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 S^2(\omega)\,d\omega \tag{I.3}$$

and hence, the product of the signal duration and the spectrum width is

$$T_e\Omega_e = \left[\int_{-\infty}^{\infty} t^2 s^2(t)\,dt \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 S^2(\omega)\,d\omega\right]^{1/2} \tag{I.4}$$

It should be borne in mind that T_e and Ω_e are root-mean-square deviations from $t = t_0$ and $\omega = 0$, respectively. The total duration of the signal should be taken at $2T_e$ and the total spectrum width (including the negative frequency range), at $2\Omega_e$.

The method of moments is inapplicable to some signals. From expressions of (I.3) it is clear that with an increase in t the time function $s(t)$ must decrease faster than $1/t$, while the function $S(\omega)$ must decrease faster than $1/\omega$, otherwise the corresponding integrals go to infinity (diverge).

In particular, this applies to the spectrum of a strictly square pulse:

$$\int_{-\infty}^{\infty} \omega^2 S^2(\omega)\,d\omega = \int_{-\infty}^{\infty} \omega^2 \frac{\sin^2\left(\frac{\omega\tau_p}{2}\right)}{\left(\frac{\omega}{2}\right)^2}\,d\omega$$

$$= 4\int_{-\infty}^{\infty} \sin^2\left(\frac{\omega\tau_p}{2}\right)d\omega = 4\int_{-\infty}^{\infty}\left(\frac{1}{2} - \frac{1}{2}\cos\omega\tau_p\right)d\omega \to \infty$$

However, for physically realizable signals the requirement of sufficiently fast decrease of $s(t)$ and $S(\omega)$ is satisfied.

Let us find the lower limit of the product $T_e\Omega_e$.

It is known that $i\omega S(\omega)$ is the spectral density of the derivative of the signal $s(t)$. Therefore, on the basis of the Parseval equation, the second integral on the right-hand side of (I.4), may be written as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 S^2(\omega)\,d\omega = \int_{-\infty}^{\infty}\left[\frac{ds(t)}{dt}\right]^2 dt$$

and then (I.4) transforms into the following expression:

$$T_e\Omega_e = \left\{\int_{-\infty}^{\infty} t^2 s^2(t)\,dt \int_{-\infty}^{\infty} [\dot{s}(t)]^2\,dt\right\}^{1/2} \tag{I.5}$$

where $\dot{s}(t) = ds/dt$.

Now let us use the Schwartz equation [see (12.11)] which, as applied to expression (I.5), assumes the form

$$T_e\Omega_e = \left| \int_{-\infty}^{\infty} t^2 s^2(t)\, dt \int_{-\infty}^{\infty} [\dot{s}(t)]^2\, dt \right|^{1/2} \geqslant \left| \int_{-\infty}^{\infty} ts(t)\, \dot{s}(t)\, dt \right| \qquad (I.6)$$

Integration by parts of the right-hand side of this inequality yields

$$\left| \int_{-\infty}^{\infty} ts(t)\, \dot{s}(t)\, dt \right| = \frac{1}{2} \left| \int_{-\infty}^{\infty} t\, \frac{d[s^2(t)]}{dt}\, dt \right|$$

$$= \frac{1}{2} \left| \int_{-\infty}^{\infty} t d_t [s^2(t)] \right| = \frac{1}{2} \left| [ts^2(t)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} s^2(t)\, dt \right|$$

The above condition that the signal $s(t)$ is quadratically integrable allows the first term on the right-hand side to be omitted. Taking into account the fact that the signal is normalized ($E = 1$), we have the following inequalities for the product $T_e\Omega_e$:

$$T_e\Omega_e \geqslant 1/2 \text{ or } T_eF_e \geqslant 1/4\pi$$

Thus, the product T_eF_e depending on the shape of the signal in any case cannot be smaller than $1/4\ \pi$.

Now let us find the signal having the minimum possible value of T_eF_e. This problem is reduced to finding the function $s(t)$ transforming inequality (I.6) into equality.

Directly from (I.6) stems the following obvious condition of equality:

$$\dot{s}(t) = cts(t)$$

where c is a constant.

Thus we obtain the equation

$$\frac{ds(t)}{dt} = cts(t)$$

or

$$\frac{ds(t)}{s(t)} = ct\, dt = \frac{1}{2} c d_t(t^2)$$

which can easily be integrated:

$$\ln s(t) = \frac{1}{2} ct^2 + b$$

Hence, the sought for function

$$s(t) = \exp\left[\frac{1}{2} ct^2 + b\right] = Ae^{ct^2/2}, \quad c < 0 \qquad (I.7)$$

is the Gaussian pulse.

[The negative sense of c results from the condition of integrability of the function $s(t)$].

In particular, for the Gaussian pulse discussed in Sec. 2.9-2 the effective duration calculated by formula (I.3) is $T_e = a/\sqrt{2}$ and the effective spectrum width $\Omega_e = b/\sqrt{2}$ (the product $T_e\Omega_e = ab/2 = 1/2$, $T_eF_e = 1/4\pi = 0.0795$).

Correspondingly, the total duration of the pulse is $2T_e = \sqrt{2}\, a$ and the spectrum width $2\Omega_e = \sqrt{2}\, b$. It should be noted that similar parameters of the Gaussian pulse determined at the level of $e^{-1/2} = 0.606$ are $2a$ and $2b$ (instead of $\sqrt{2}a$ and $\sqrt{2}b$).

For a triangular pulse whose spectral density (Fig. I.1*a* and *b*) is

$$S(\omega)=\frac{A\tau_p}{2}\left(\frac{\sin\frac{\omega\tau_p}{4}}{\frac{\omega\tau_p}{4}}\right)^2$$

the parameters T_e and F_e are as follows:

$$T_e=\frac{1}{\sqrt{40}}\tau_p,\quad F_e=\frac{\sqrt{12}}{2\pi\tau_p},\quad T_eF_e=0.0873,\quad \Omega_eT_e=\sqrt{12/40}=0.55$$

We may see that the magnitude of the product of duration and frequency band for a triangular pulse is only 10% greater than that for the Gaussian pulse.

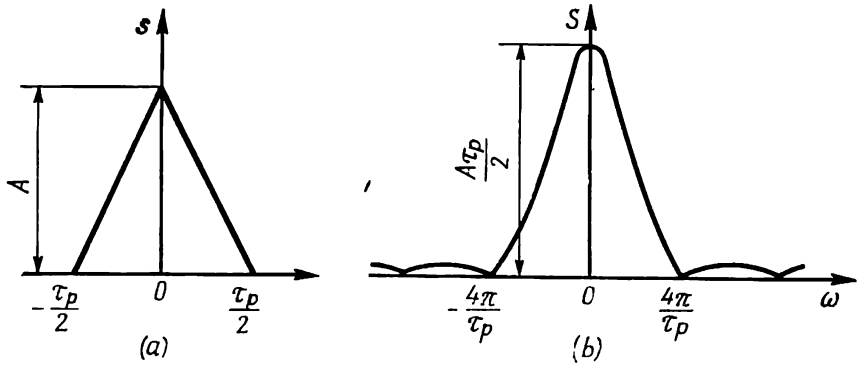


Fig. I.1. (*a*) Triangular pulse and (*b*) its spectral density

From the above relations it follows that compression of a pulse in time, e.g., for increasing the accuracy of measurement of the moment of its appearance, is always accompanied by the widening of the pulse spectrum, and this makes it necessary to broaden the passband of the measuring system. In a similar way, compression of a pulse for increasing the accuracy of frequency measurements is always accompanied by the expansion of the signal in time, and this requires that the time of measurement be increased. The impossibility of simultaneous concentration of a signal in a narrow frequency band and in a short interval of time is one of the manifestations of the uncertainty principle known in physics.

APPENDIX II

AUTOCORRELATION FUNCTION OF A SIGNAL ON THE TIME-FREQUENCY PLANE

In the preceding chapters the autocorrelation function was used for evaluating the degree of connection between the signal $s(t)$ and its copy $s(t+\tau)$ shifted *in time* by an amount τ.

A similar problem arises when defining the shift of the spectrum of a narrow-band signal along the *frequency axis.* Therefore, it is expedient to generalize the concept of the autocorrelation function to cover the case of a simultaneous shift of the signal both in time and in frequency.

Let us write the original oscillation in the form

$$a(t) = A(t) \cos[\omega_0 t + \theta(t)]$$

and the same oscillation shifted in time by an amount τ and in frequency by an amount $\Omega = 2\pi F$ as

$$a_{\tau,\,\Omega}(t) = A(t+\tau) \cos[(\omega_0 + \Omega)(t+\tau) + \theta(t+\tau)]$$

Transforming to analytic signals, we obtain respectively

$$z(t) = A(t)\, e^{i\theta(t)} e^{i\omega_0 t} = \mathbf{A}(t)\, e^{i\omega_0 t}$$

$$z_{\tau,\,\Omega}(t) = A(t+\tau)\, e^{i\theta(t+\tau)} e^{i(\omega_0+\Omega)(t+\tau)} = \mathbf{A}(t+\tau)\, e^{i(\omega_0+\Omega)\tau} e^{i(\omega_0+\Omega)t} \qquad \text{(II.1)}$$

Then the autocorrelation function of the analytic signal can be defined by the expression [see (3.94) and (3.97)]

$$B_z(\tau,\ \Omega) = \int_{-\infty}^{\infty} z(t)\, z^*_{\tau,\,\Omega}(t)\, dt = e^{-i(\omega_0+\Omega)\tau} \int_{-\infty}^{\infty} \mathbf{A}(t)\, \mathbf{A}^*(t+\tau)\, e^{-i\Omega t}\, dt \qquad \text{(II.2)}$$

The expression thus obtained takes into account both the shift of the signal $a(t)$ in time (τ) and the shift in frequency ($\Omega = 2\pi F$) and is called the *generalized* autocorrelation function.

The modulus of expression (II.2)

$$|B_z(\tau,\ \Omega)| = \left| \int_{-\infty}^{\infty} \mathbf{A}(t)\, \mathbf{A}^*(t+\tau)\, e^{-i\Omega t}\, dt \right| \qquad \text{(II.3)}$$

is called the *two-dimensional* autocorrelation function.

Widely used in the theory of signals is the following normalized two-dimensional autocorrelation function:

$$\rho(\tau,\ \Omega) = \frac{|B_z(\tau,\ \Omega)|}{B_z(0,\ 0)} = \frac{\left| \int_{-\infty}^{\infty} \mathbf{A}(t)\, \mathbf{A}^*(t+\tau)\, e^{-i\Omega t}\, dt \right|}{\int_{-\infty}^{\infty} A^2(t)\, dt} \qquad \text{(II.4)}$$

In the rectangular system of coordinates τ, Ω, ρ, the function $\rho(\tau, \Omega)$ is shown in the form of a surface, an example of which for a Gaussian radio pulse (with an unmodulated r-f carrier) is given in Fig. II.1.

Directly from expression (II.4) it follows that the maximum value of the function $\rho(\tau, \Omega)$, i.e., $\rho(0, 0)$, is equal to unity.

It has been found that the volume of the body limited by the plane $\rho = 0$ and the surface $\rho^2(\tau, \Omega)$ is also equal to unity *regardless of the laws governing the modulation of the amplitude and phase of the signal* [1], [13]:

$$V_{\rho^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty}\int \rho^2(\tau,\ \Omega)\, d\tau\, d\Omega = 1 \qquad \text{(II.5)}$$

If, by changing modulation, one compresses the body under the surface $\rho^2(\tau, \Omega)$ along the axis τ, it spreads along the axis Ω; compression along the axis Ω results in extension along the axis τ.

This shows that a simultaneous increase of the signal resolution both in time and in frequency is impossible, and relation (II.5) is a mathematical wording of the uncertainty principle in radar.

In this connection, the two-dimensional autocorrelation function $\rho(\tau, \Omega)$ of a signal is often called the *uncertainty function*, while the body limited by the plane $\rho = 0$ and the surface $\rho^2(\tau, \Omega)$ is referred to as the *uncertainty body* (Fig. II.1).

Setting in (II.2) $\Omega = 0$, we obtain the expression

$$B_z(\tau, 0) = e^{-i\omega_0\tau} \int_{-\infty}^{\infty} \mathbf{A}(t)\, \mathbf{A}^*(t+\tau)\, dt \qquad \text{(II.6)}$$

coinciding with (3.97).

To transform to the autocorrelation function $B_a(\tau)$ of the physical oscillation $a(t) = A(t)\cos[\omega_0 t + \theta(t)]$, one should separate the real part of $B_z(\tau, 0)$ in accordance with (3.96). The integral in expression (II.6) defines the *complex envelope of the function* $B_z(\tau, 0)$, while the rapidly oscillating factor $e^{-i\omega_0\tau}$ defines the r-f carrier of this function.

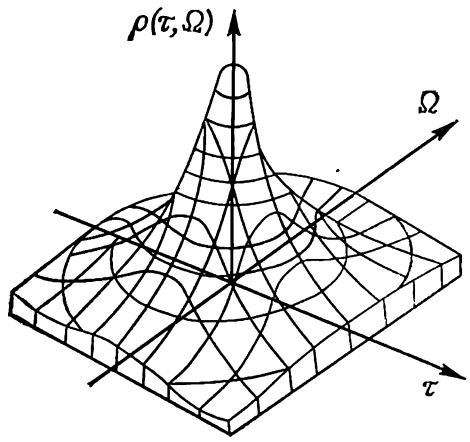


Fig. II.1. Uncertainty body of a Gaussian radio pulse

Transforming to the modulus and taking into account that $|e^{-i\omega_0\tau}| = 1$, we obtain

$$|B_z(\tau, 0)| = \left| \int_{-\infty}^{\infty} \mathbf{A}(t)\, \mathbf{A}^*(t+\tau)\, dt \right| \qquad \text{(II.7)}$$

From this expression it follows that the section of the surface of the uncertainty body by the plane $\Omega = 0$ defines *the envelope of the autocorrelation function* $B_a(\tau)$ *of the oscillation*

$$a(t) = A(t)\cos[\omega_0 t + \theta(t)]$$

On the other hand, setting in (II.3) $\tau = 0$, we get the expression

$$|B_z(0, \Omega)| = \left| \int_{-\infty}^{\infty} A^2(t)\, e^{-i\Omega t}\, dt \right|$$

which, on the basis of expressions (2.61) and (2.62), can be reduced to the form

$$|B_z(0, \Omega)| = \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}_A(x)\, \mathbf{S}_A(\Omega - x)\, dx \right|$$

$$= \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{S}_A(x)\, \mathbf{S}_A^*(x - \Omega)\, dx \right| \qquad \text{(II.8)}$$

where $\mathbf{S}_A(\omega)$ is the spectral density of the envelope $\mathbf{A}(t)$.

Thus, the section of the surface of the uncertainty body by the plane $\tau = 0$ defines the function of frequency autocorrelation, i.e., the degree of connection of the spectrum $S_A(\omega)$ with its copy shifted along the frequency axis by an amount Ω.

The two-dimensional autocorrelation function $\rho(t, \Omega)$ defined by expression (II.4) acquires special importance in the *matched filtering* of a signal. In Ch. 12 it was shown that the signal at the output of a matched filter coincides with the autocorrelation function of the input signal. Therefore, the autocorrelation function $|B_z(\tau, 0)|$ can be used for evaluating the resolution of the signal in time, while the function $|B_z(0, \Omega)|$, in frequency.

These properties of the uncertainty function form the basis of the theory of resolution of radar systems in determining the range of the target and its speed.

LITERATURE

To Chaps. 1 through 4

1. Трахтман А. М. Введение в обобщенную спектральную теорию сигналов. М., «Сов. радио», 1972. (Trakhtman A. M. An introduction to the generalized spectrum theory of signals. Moscow, «Soviet Radio» Publishers, 1972).
2. Зиновьев А. Л., Филиппов Л. И. Введение в теорию сигналов и цепей. 2-ое Изд., М., «Высшая школа», 1975. (Zinovyev A. L., Filippov L. I. An introduction to the theory of signals and circuits. Second Ed. Moscow, «Vysshaya Shkola» Publishers, 1975).
3. Варакин Л. Е. Теория сложных сигналов. М. «Сов. радио», 1970. (Varakin L. E. Theory of complex signals, Moscow, «Soviet Radio» Publishers, 1970).
4. Franks L. E. Signal theory. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1969.
5. Конторович М. И. Операционное исчисление и процессы в электрических цепях. М., «Сов. радио», 1975. (Kontorovich M. I. Operational calculus and processes in electric circuits. Moscow, «Soviet Radio» Publishers, 1975).
6. Гоноровский И. С. Радиосигналы и переходные явления в радиоцепях. М., Связьиздат, 1954. (Gonorovsky I. S. Radio Signals and Transient Phenomena in Radio Circuits. Moscow, «Svyazizdat» Publishers, 1954).
7. Левин Б. Р. Теоретические основы статистической радиотехники. М., «Сов. радио», 1965. (Levin B. R. Theoretical fundamentals of statistical radio engineering. Moscow, «Soviet Radio» Publishers, 1966).
8. Тихонов В. И. Статистическая радиотехника. М., «Сов. радио», 1966. (Tikhonov V. I. Statistical radio engineering. Moscow, «Soviet Radio» Publishers, 1966).
9. Davenport W. B. and Root W. L. An introduction to the theory of random signals and noise. McGraw-Hill Book Company, Inc., New York, Toronto, London, 1958.
10. Гоноровский И. С., Фельдман Л. Д., Васильев А. В. О спектрах некоторых неинтегрируемых функций. «Радиотехника и электроника», 1970, № 4. (Gonorovsky I. S., Feldman L. D., Vasiliev A. V. On spectra of some nonintegrable functions. J. «Radiotekhnika i Elektronika», 1970, No. 4).
11. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М., ГИФМЛ, 1971 (Gradshtein I. S., Ryzhik I. M. Tables of integrals, sums, series and products. Moscow, GIFML Publishers, 1971).

To Chaps. 5 through 7

1. Степаненко И. П. Основы теории транзисторов и транзисторных схем. М., «Энергия», 1973. (Stepanenko I. P. Fundamentals of the theory of transistors and transistor circuits, Moscow, «Energiya» Publishers, 1973).
2. Атабеков Г. И. Основы теории цепей. М., «Энергия», 1969. (Atabekov G. I. Fundamentals of the theory of circuits, Moscow, «Energiya» Publishers, 1969).

3. Karni S. Networks theory: analysis and synthesis. Allyn and Bacon, Inc., Boston, Massachusetts, 1966.
4. Конторович М. И. Операционное исчисление и процессы в электрических цепях. М., «Сов. радио», 1975. (Kontorovich M. I. Operational calculus and processes in electric circuits. Moscow, «Soviet Radio» Publishers, 1975).
5. Гоноровский И. С. Радиосигналы и переходные явления в радиоцепях. М., Связьиздат, 1954. (Gonorovsky I. S. Radio signals and transient phenomena in radio circuits. Moscow, «Svyazizdat» Publishers, 1954).
6. Левин Б. Р. Теоретические основы статистической радиотехники. М., «Сов. радио», 1966. (Levin B. R. Theoretical fundamentals of statistical radio engineering. Moscow, «Soviet Radio» Publishers, 1966).
7. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М., ГИФМЛ, 1971. (Gradshtein I. S., Ryzhik I. M. Tables of integrals, sums, series and products, Moscow, GIFML Publishers, 1971).

To Chaps. 8 through 10

1. Зиновьев А. Л., Филиппов Л. Н. Введение в теорию сигналов и цепей. М., «Высшая школа», 1975. (Zinovyev A. L., Filippov L. N. An introduction to the theory of signals and circuits. Second ed. Moscow, «Vysshaya Shkola» Publishers, 1975).
2. Кушнир В. Ф., Ферсман Б. А. Теория нелинейных электрических цепей. М., «Связь», 1974. (Kushnir V. F., Fersman B. A. Theory of nonlinear electric circuits. Moscow, «Svyaz» Publishers, 1974).
3. Андреев В. С. Теория нелинейных электрических цепей. М., «Связь», 1972. (Andreev V. S. Theory of nonlinear electric circuits, Moscow, «Svyaz» Publishers, 1972).
4. Крылов Н. Н. Электрические процессы в нелинейных элементах радиоприемников. Связьиздат, 1949. (Krylov N. N. Electrical processes in nonlinear elements of radio receivers. Moscow, «Svyazizdat» Publishers, 1949.
5. Кобзарев Ю. Б. О нелинейном методе трактовки явлений в ламповом генераторе. «ЖТФ», 1935, № 5. (Kobzarev Yu. B. On nonlinear method of treatment of phenomena in a tube oscillator. J. «ZH.T.F.», 1935, No. 5).
6. Боголюбов Н. Н., Митропольский Ю. А. Асимптотические методы в теории нелинейных колебаний. М., Физматгиз, 1958. (Bogolyubov N. N., Mitropolsky Yu. A. Asymptotic methods in the theory of nonlinear oscillations. Moscow «Fizmatgiz» Publishers, 1958).
7. Гоноровский И. С. К вопросу об установлении автоколебаний в высокочастотном генераторе с запаздывающей обратной связью. «Радиотехника», 1958, № 5. (Gonorovsky I. S. On origination of self-oscillations in a radiofrequency oscillator with delayed feedback. J. «Radiotekhnika», 1958, No. 5).

To Chap. 11

1. Левин Б. Р. Теоретические основы статистической радиотехники. М., «Сов. радио», 1966. (Levin B. R. Theoretical fundamentals of statistical radio engineering. Moscow, «Soviet Radio» Publishers, 1966).
2. Тихонов В. И. Статистическая радиотехника. М., «Сов. радио», 1966. (Tikhonov V. I. Statistical radio engineering. Moscow, «Soviet Radio» Publishers, 1966).
3. Davenport W. B. and Root W. L. An introduction to the theory of random signals and noise. McGraw-Hill Book Company, Inc., New York, Toronto, London, 1958.
4. Кремер И. Я., Владимиров В. И., Карпухин В. И. Модулирующие помехи и прием радиосигналов. М., «Сов. радио», 1972. (Kremer I. Ya., Vladimi-

rov V. I., Karpukhin V. I. Modulating Noise and Reception of Radio Signals. Moscow, «Soviet Radio» Publishers, 1972).
5. Смирнов В. И. Курс высшей математики. Т. 3, ч. 2. М., «Наука», 1970. (Smirnov V. I. Course of higher mathematics. Vol. 3, part 2. Moscow, «Nauka» Publishers, 1970).
6. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М., ГИФМЛ, 1971. (Gradshtein I. S., Ryzhik I. M. Tables of integrals, sums, series and products. Moscow, GIFML Publishers, 1971).

To Chaps 12 through 15 and Appendices

1. Котельников В. А. Теория потенциальной помехоустойчивости М.-Л., Энергоиздат, 1956. (Kotelnikov V. A. Theory of potential noise immunity. Moscow-Leningrad, Energoizdat Publishers, 1956).
2. Ширман Я. Д. Разрешение и сжатие сигналов. М., «Сов. радио», 1974. (Shirman Ya. D. Resolution and compression of signals. Moscow, «Soviet Radio» Publishers, 1974).
3. Gold B. and Rader M. Digital processing of signals. McGraw-Hill Book Company, Inc., New York, Toronto, London, 1969.
4. Трахтман А. М., Трахтман В. А. Основы теории сигналов на конечных интервалах. М., «Сов. радио», 1975. (Trakhtman A. M., Trakhtman V. A. Fundamentals of the theory of signals on finite intervals. Moscow, «Soviet Radio» Publishers, 1975).
5. Трахтман А. М. Введение в обобщенную спектральную теорию сигналов. М., «Сов. радио», 1972. (Trakhtman A. M. An introduction to the generalized spectrum theory of signals. Moscow, «Soviet Radio» Publishers, 1972).
6. Фоменко И. Б. Анализ случайных процессов с использованием функций Уолша. «Радиотехника и электроника», 1977, № 4. (Fomenko I. B. Analysis of random processes with the use of Walsh functions. J. «Radiotekhnika i Elektronika», 1977, No. 4).
7. Guillemin E. A. Synthesis of passive networks. Wiley J. and sons, Inc., New York, 1957.
8. Karni S. Networks theory: analysis and synthesis. Allyn and Bacon, Inc., Boston, Massachusetts, 1966.
9. Матханов П. Н. Основы синтеза линейных электрических цепей. М., «Высшая школа», 1976. (Matkhanov P. N. Fundamentals of synthesis of linear electric circuits. Moscow, «Vysshaya Shkola» Publishers, 1976).
10. Huelsman A. P. Theory and design of active RC circuits. McGraw-Hill Book Company, Inc., New York, Toronto, London, 1968.
11. Знаменский А. Е., Теплюк И. Н. Активные RC-фильтры. М., «Связь», 1970. (Znamensky A. E., Teplyuk I. N. Active RC-filters. Moscow, «Svyaz» Publishers, 1970).
12. Hansel G. E. Filter design and evaluation. Van Nostrand Reinhold Company, 1969.
13. Bode H. W. Network analysis and feedback amplifier design. New-York, 1946.
14. Franks L. E. Signal theory. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1969.
15. Yanke E., Emde F., Lösch F. Tafeln höherer Funktionen. B. G. Teubner Verlagsgesselschaft, Stuttgart, 1960.
16. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М., ГИФМЛ, 1971 (Gradstein I. S., Ryzhik I. M. Tables of integrals, sums, series and products. Moscow, GIFML Publishers, 1971).

LIST OF SYMBOLS

A	amplitude
A_n	amplitude of nth harmonic
A_0	carrier amplitude
$A\ (t)$	envelope of r-f oscillation
$\dot{A}\ (t)$	derivative of function $A\ (t)$
$A_c(t),\ A_s\ (t)$	quadrature components of envelope $\mathbf{A}\ (t)$
$\mathbf{A}\ (t)$	complex envelope
$a,\ a_{eqv}$	detuning parameter of circuit
a_n	cosine Fourier series coefficient
$B_s\ (\tau)$	autocorrelation function of oscillation $s\ (t)$
$B_z\ (\tau)$	autocorrelation function of analytic signal $z\ (t)$
$B_A\ (\tau)$	autocorrelation function of envelope $A\ (\tau)$
$B_{s_1 s_2}\ (\tau)$	cross-correlation function of signals $s_1\ (t)$ and $s_2\ (t)$
$B_x\ (\tau)$	autocorrelation function of random process
$B_{xy}\ (\tau)$	cross-correlation function of random processes $x\ (t)$ and $y\ (t)$
b_n	sine Fourier series coefficient
C	capacitance
c_n	generalized Fourier series coefficient
E	e. m. f. amplitude
E_0	carrier e. m. f. amplitude
E_s	signal energy dissipated in resistor of 1Ω
$E\ (t)$	envelope of radio-frequency e.m.f. $e\ (t)$
$\mathbf{E}\ (t)$	complex envelope of radio-frequency e.m.f.
$\mathbf{E}\ (\omega)$	spectral density of e.m.f.
$e\ (t)$	instantaneous e.m.f.
e_p	pumping voltage
$f,\ F$	frequency
f_0	centre (carrier) frequency
f_m	maximum frequency
G_i	internal conductance of signal source
G_l	load conductance
$G\ (t)$	impulse response envelope
$g\ (t)$	instantaneous impulse response
I	current amplitude
$\mathbf{I}$	complex current amplitude
$i\ (t)$	instantaneous current
$J_n\ (m)$	Bessel function
$\mathbf{K}\ (i\omega),\ \mathbf{K}\ (p)$	complex transfer function
$\mathbf{K}\ (\omega)$	transfer function modulus
$\hat{K}\ (z)$	transfer function of discrete filter
$\mathbf{K}_0\ (i\omega),\ \mathbf{K}_0\ (p)$	transfer function of circuit with feedback
$\mathbf{K}_T\ (i\omega),\ \mathbf{K}_T\ (p)$	transfer function of discrete filter
$\mathbf{K}_a\ (i\omega),\ \mathbf{K}_a\ (p)$	transfer function of amplifier
k_{am}	slope of characteristic curve of amplitude modulator
k_{fm}	slope of characteristic curve of frequency modulator
k_{pm}	slope of characteristic curve of phase modulator
L	inductance
$L_s\ (p)$	Laplace transform of function $s\ (t)$
M	modulation factor, mutual inductance
m	angle modulation index, base of linear FM pulse
p	complex variable
$p\ (x)$	univariate probability density
Q	Q-factor of oscillatory circuit
$R_x\ (\tau)$	normalized autocorrelation function
S	slope of characteristic curve of active element

$\mathbf{S}(\Omega)$, $\mathbf{S}(\omega)$	complex spectral density of function $s(t)$
$s(kT)$	signal sample at instant $t = kT$
$S(\Omega)$, $S(\omega)$	spectral density modulus
$\mathbf{S}_A$	spectral density of envelope $\mathbf{A}(t)$
$s(t)$	instantaneous value of signal
$\overline{s(t)}$	time average of $s(t)$
$\hat{\mathbf{S}}(z)$	z-transform
$s_T(t)$	signal $s(t)$ sampled at intervals T
$\mathbf{S}_T(\omega)$	spectral density of discrete signal
$\mathbf{S}_T(n\ \omega_1)$	discrete Fourier transform
$\mathbf{S}_T(p)$	discrete Laplace transform
$\langle s(t) \rangle$	ensemble average of $s(t)$
T	oscillation period
T_s	signal duration
V	voltage amplitude
$V(t)$	envelope of r-f oscillation $v(t)$
$\mathbf{V}(t)$	complex envelope of voltage $v(t)$
$\mathbf{V}(\omega)$	spectral density of voltage
$v(t)$	instantaneous voltage
W_0	energy spectrum of white noise
$W_x(\omega)$	energy spectrum of random process $x(t)$
$W_A(\Omega)$	energy spectrum of random process envelope
$W_{xy}(\omega)$	mutual energy spectrum of random processes $x(t)$ and $y(t)$
Y	admittance
Z	impedance
$Z(\omega)$	spectral density of complex oscillation $z(t)$
$z_a(t)$	analytic signal corresponding to physical signal $a(t)$
α	attenuation, physical parameter of transistor (see Ch. 5)
β	rate of change of angular frequency, physical parameter of transistor
γ	modulating function phase (in AM — envelope phase)
γ_g	phase of r-f carrier of impulse response
Δt	sampling interval
$\Delta\omega$, Δf	frequency quantization step, detuning
$\Delta\omega_0$, Δf_0	half-width of narrow-band spectrum, passband of filter
$\delta(t)$, $\delta(\omega)$	delta function
θ	current cutoff angle
θ_n	epoch angle of nth harmonic
$\theta(\omega)$	argument of complex spectral density of signal
θ_0	epoch angle of carrier wave
$\theta(t)$	instantaneous value of phase of narrow-band oscillation
θ_{max}	maximum phase variation in angle modulation
ρ	characteristic impedance of oscillatory circuit
$\rho_x(\tau)$	covariance function of random process
σ_x^2	variance of random process
τ	time
τ_p	pulse duration (length)
τ_0, τ_1, τ_a, τ_k	time constants of respective circuits
φ	argument of transfer function of circuit
$\varphi(\omega)$	phase-frequency characteristic of two-port network
$\varphi_n(t)$	basic function of orthogonal set
$\psi(t)$	phase angle of radio-frequency oscillation
ω, Ω	angular frequency
ω_m	maximum frequency
ω_d	frequency deviation
ω_p	pumping frequency

INDEX

Addition, modulo 2, 562
  of oscillations, 55
Admittance, forward-transfer, 188
  input, 188
  output, 188
    open circuit, 199
  reverse-transfer, 188
Amplification, parametric, 408-409, 414-422, 423
Amplifier, aperiodic, 204-207
    discrete signals in, 236-239
  constant-voltage current, 201
  electron-tube, 202-204
  linear, 192-195
  operational, 593
  parametric, biharmonic operation of, 416
    double-tuned, 416-421
    single-tuned, 414-416
    synchronous operation of, 415
  regenerative, 373-376
  resistance-coupled, 204
    intrinsic noise in, 280-281
  resonance, 207-210
    intrinsic noise in, 281-284
    nonlinear, 309-313, 342-344
    radio pulse in, 250-256
  transistor, 195-202
Amplitude, complex, 133
Angle, conduction (operating), 304
  current cutoff, 304
Approximation, Butterworth, 601
  Chebyshev, 605
  maximum flat, 601
  of characteristic, 299
  piecewise-linear, 301
  polynomial, 299
Arrangement, commom-base circuit, 196-198, 201
  commom-collector circuit, 198, 201
  common-emitter circuit, 198, 199-200, 201
Attenuation, 251, 379
Averaging, ensemble, 147
  time, 149

Base, of signal, 88
Body, uncertainty, 627

Capacitance, differential, 298, 389
  parametric, 389
  periodically varying, 410-413, 423-431
Channel, communication, 29
Change, of time scale, 53
Characteristic, amplitude-frequency, 50, 72
  current-voltage, 24, 27-28
    slope of, 24-25
  detection, 329
  normalized frequency, 221
  oscillatory, 352
  phase-frequency, 50, 72
  spectral, of function, 49
  volt-ampere, 297
  volt-coulomb, 297
  volt-farad, 297
Circuit(s), active, 185
  automatic bias-control, 352, 353
  bridge, 594-598
    input admittance of, 597
  classification of by the types of signal, 22
  differentiating, 239
  idle tank, 416
  impedance-divider oscillator, 358-359
  integrating, 239
  linear, classification of, 23
    with constant parameters, 26
    with variable parameters, 26
    parametric, 28, 388, 410
  nonlinear, 23, 27, 28
  operating tank, 416
  phase-correcting, 591, 594-598
Code, parallel, 501
Coefficient, compression, 485
  correlation, 439
  coupling, 335
  Fourier, 37
  weighting, 503, 505
Coherence, 95
Component(s), quadrature, 114
Condition(s), for compensation of epoch angles in signal spectrum, 469
  for self-excitation, 350-351
    hard, 356
    soft, 356
  of orthogonality, 37
  of orthonormality, 548
Conjugate, harmonic, 128
Constant, Euler, 464
  time, of *R-C* circuit, 238
Content, harmonic, 273, 326
Conversion, frequency, 337-341
  by means of parametric amplifier, 422-423
  spectrum inversion in, 340-341
Converter, analog-to-digital, 500, 536-541
  digital-to-analog, 501, 541-545
Convolution, of functions, 236

Correlation, between signals, 90
Criterion, Paley-Wiener, 474
  stability, 223
    algebraic, 225
    Nyquist, 229
    Routh-Hurwitz, 225
Current, electron, 277-279
  quiescent, 325
Curve, cutoff cosine, 304

Decoding, 501
Degree of coherence, 96
  of modulation, 100
Delay of message, 260
Demodulation (detection), 15, 324
Density, average power spectral, 159
  cross-spectral, 167
  joint probability, 171
  one-dimensional (univariate) probability, 146
  spectral, of function, 49
    of oscillation energy, 58
  two-dimensional (bivariate) probability, 147
Detection, 20, 324
  linear, 441-443, 446-447, 449, 454
  of low amplitudes, 325
  square-law, 443-445, 447-448, 449, 454
  synchronous, 341
Detector, amplitude, 324-331, 441-449
    linear, 328
    square-law, 325
  frequency, 331-337, 449-452
    double-tuned, 333-337
    single-tuned, 332
  phase, 337, 496
Determinant, Wronskian, 400
Detuning, 208, 209, 246
Deviation, frequency, 111
    instantaneous, 331
  mean-square, 147
Differentiation of oscillation, 54
Distortion, linear frequency, 259, 269
  nonlinear, 261, 269, 329, 375 :
Distribution, normal (Gaussian), 62, 155
  probability, 172, 441
  spectral power, 33
Duration, signal, 622

Effect, shot, 277
Electromagnetic waves, subdivision of into bands, 15
Element, feedback, 211
  nonlinear reactive, 296
    resistive, 296, 302
Energy, interaction, 96
Envelope, complex, 134, 495
    autocorrelation function of, 137
    spectral density of, 135-136
  normalized, 254
Equation, Mathieu, 424
  Parseval's, 58
  phase-balance, 354-355
  Van der Pol's, 363
Ergodicity, 149
Error, mean-square, 35
  quantization, 537
  sampling, 510
Excitation, biharmonic, 401-407
  parametric, 409, 431
Expectation, mathematical, 147

Factor, amplification, 203, 312
  convergence, 72
  feedback, 347
  modulation, 100, 101
  peak, 33, 538
  penetration, 202
  reciprocal duty, 46
Fading, 16
Feedback, conductive (autotransformer), 358
  current, 212
  external, 211, 346
  inductive (transformer), 358
  internal, 211, 368
  negative, 213
  positive, 213, 345, 373
  voltage, 212
Filter, comb, 488
    rejection, 519
  complex, 497-498
  digital, 501
    speed of response of, 545, 547
    synthesis of, 616-621
  discrete, first-order, 517-519
    recoursive, 519-522, 531-533
    second-order nonrecoursive, 530-531
    recoursive, 533-536
    transfer function of, 515-522
  Gaussian, 475
  high-pass, 608-609
    Butterworth, 609
    Chebyshev, 609
  inverted, 489
  low-pass, 600-608
    Butterworth, 601-605
      transfer function of, 603-604

Chebyshev, 605-608
transfer function of, 606-608
matched, 466-488
amplitude characteristic of, 471, 475
phase characteristic of, 471
physical realizability of, 474-475
synthesizing, 501, 544-545
Filtering, linear, 467
matched, 467, 468-472, 628
Follower, emitter, 201
Frequency, angular, 110
mirror, 319
side, 103
Function(s), autocorrelation, 90
generalized, 626
normalized, 150
of modulated wave, 138-141
two-dimensional, 626, 628
autocovariance, 150
Berg, 306
Butterworth, 601
characteristic, 291
cross-correlation, 93, 165
delta, 68
filtering property of, 69
Euler's, 464
Green's, 401
Hermite, 554
Laguerre, 551-553
Mathieu's, 424
modulating, 98
needle, 76
normalized, 34
Rademacher, 559-560
Rayleigh, 445
sampling, 85
sign, 559-560
slowly-varying, 99
transfer, 56, 214
uncertainty, 627
unit-impulse, 68
Walsh, 560-577
discrete, 577-581
Hadamard-ordered, 564-565
Paley-ordered, 563
Walsh-ordered, 562
weighting, 233, 548

Gain, current, 192
power, 192
voltage, 192
Gate, coincidence, 568
Generator, Laguerre function, 552-553
Gyrator, 612-616

Harmonics, 28
Hodograph, of function, 228

Impedance, characteristic, 369
input, 192-193
output, 193
resonant, 369
Impulse, unit, 68
Index, angle modulation, 111
Inductance, differential, 298
parametric, 390
periodically-varying, 413-414
Inequality, Bessel's, 36
Schwarz, 470
Integral, double Fourier, 49
Fresnel, 481
probability, 156
sine, 512
Integration, of oscillation, 55
Integrator, ideal, 244
Interval, correlation, 95
sampling, 510

Keying, frequency-shift, 263-269
parasitic amplitude modulation in, 268
phase-shift, 261-263
parasitic amplitude modulation in, 263

Lemma, Riemann's, 67
Limiter, amplitude, 315-320, 453-457
Limiting, amplitude, 386-387
Line, feedback, 353
surface-wave dispersive, 482
Link, second-order, 591-594
Linkage, magnetic, 298

Margin, of circuit stability, 610
Matrix, admittance, 186
coherency, 165
Hadamard, 564
hybrid, 187
impedance, 186
Meander, 42
Message, complex, 259
Method, approximate spectral, 246-249
envelope, 249-250
of characteristic functions, 291-295
of moments, 622-623
quasi-linear, 312
spectral, 232-234
superposition integral, 234-236
zero beat, 339
z-transform, 523-524
Mixer, 340
Modulation, 14, 19

amplitude, 99, 100-109, 177, 181
linear distortion in, 256-261
parasitic, 126, 163, 168, 273-274, 382, 396
angle, 99, 109-123, 380
fast, 119
slow, 119
as a parametric process, 394-396
bias, 344
combined amplitude-frequency, 123-126
frequency, 99, 111, 380-382
percentage, 100
phase, 99, 111, 181-184
parasitic, 104
Moment, first, 147
second, 147
second mixed, 148
Multiplier, frequency, 313-315

Network(s), three-terminal, 195
two-port, 185-191, 583-616
logarithmic attenuation of, 585
minimum-phase, 588
nonminimum-phase, 588
transfer function of, 193, 583-584
transmission constant of, 585
Noise, additive, 434
band-limited, 163
external, 29
internal, 30, 277-284
multiplicative, 434, 461-465
narrow-band, 163
nonwhite, 492
quantization, 538, 540-541
rounding-off, 545
thermal, 284
white, 161
wide-band, 161
Norm, of function, 34
Number, of degrees of freedom, 88

Oscillation(s), 33
amplitude-modulated, generation of, 341-344
average power of, 48
carrier, 98
biharmonic, 307
combination, 308
frequency-modulated, 269-274
frequency-shift keyed, 264
harmonic, 71
of random amplitude, 152
of random phase, 153-155
modulated, 99
phase-shift keyed, 262
radio-frequency, 98
generation of, 19
sampling of, 142
saw-tooth, 44
self-excited, 28
square-wave, 42
Oscillator, parametric, 431-433
amplitude limiting in, 432
pumping, 401
reference, 501

Passband, relative, 210
Prametron, 431-433
Phenomenon, Gibbs, 44
Polynomial(s), Chebyshev, 549-551, 558-559
Hermite, 553-554
Hurwitz, 583
Laguerre, 551
Legendre, 549, 556-557
Principle, superposition, 25
uncertainty, 627
Probability, integral, 146
Problem, detection, 440
Process, narrow-band, 98, 127
quadrature, 496
Product, of two oscillations, 55
Property, group, 580
Pulling, frequency, 376
Pulse, clock, 502
exponential, 72
Gaussian, 62, 624
half-wave cosine, 555-559
linear FM, 120, 481-485, 494-495, 499
of sinc(x) form, 64
of unit area, 68
rectangular, 68
square, 59-60
video, 477-480
triangular, 625

Quantization, level, 500
time, 500, 539-540

Random function, differentiation of, 284-287
integration of, 284-287
Random process, autocorrelation function of, 148
average power of, 278
delta correlated, 161
ergodic, 149
narrow-band, 168-176
envelope of, 169-173
frequency of, 176
phase of, 175-176
normal, 155-158

normalization of in narrow-band linear circuits, 289-291
strictly stationary, 148
wide-sense stationary, 148
Range, pulling, 379
Ratio, open-circuit reverse voltage transfer, 199
pulse period-to-pulse width, 46
rejection, 319
signal-to-noise, 446, 448-449, 452, 456-457, 477, 493, 538
maximization of, 470-472
transfer, 194
Rectifier, half-wave, 320-324
Regeneration, 373
Resistance, differential, 195, 368
negative, 349, 367, 369
Response, amplitude-frequency, 194, 209
impulse, 56, 194
of ideal differentiator, 245
of ideal integrator, 244
of matched filter, 473-474
of parametric circuit, 396-401
phase-frequency, 194, 209
transient, 194

Sample, complex, 89
frequency, 88, 507
time, 88
Sampling, 500, 502, 539-540
Selectivity, of receiver, 20
Self-excitation, 345
Self-oscillator, 345-387
built around tunnel diode, 369-370
dynatron, 368
local, 360-461
R-C, 382-388
nonlinear equation of, 361-367
single-tuned transistor, 352
common-base, 360-361
single-tuned vacuum-tube, 348-351
Colpitts (tapped-capacitor), 358-359
Hartley (tapped-coil), 358-359
with delayd feedback, 370-373
with internal feedback, 367-370
Sensitivity, pole, 610
Series, Fourier trigonometric, 39
Fourier complex, 39
generalized Fourier, 35
Set(s), orthogonal, of functions, 34
complete, 36
orthonormal, of functions, 34
Shift, arithmetic, 580
dyadic, 579-581
of oscillation spectrum, 54
phase, of message, 272-273
time, of oscillation, 52
Signal(s), analytic, 133-138
autocorrelation function of, 136-138
energy of, 137
spectrum of, 134-135
classification of, 21
deterministic, 32
narrow-band, sampling of, 141-145
periodic, 32
phase-shift keyed, 490
nonperiodic, 33
random, 33
random-level d-c voltage, 151
Slope, of characteristic curve, 199, 202, 296-297, 331
average, 297, 311
differential 297, 311, 388
Source, dependent, 188
current-controlled current, 202
voltage-controlled current, 201
voltage-conrrolled voltage, 201
Spectrum, continuous, 49
energy, 159
of frequencies, 26
signal, transformation (conversion) of, 28
Stability, relative frequency, 19
Stationarity, of random process, 149
Step, unit, 71
spectrum of, 74
quantizing, 537
System, key-lock, 492
Symmetry, quadrantal, 599

Table, truth, 568
Theorem, central limit, 277, 291
convolution, 509-510
delay, 509
Hurwitz, 224-225
Kotelnikov sampling, 84
Manly-Raw, 406
reciprocity, 189
Wiener-Khintchine, 160
Threshold, limitation, 450
Train, of identical pulses, 65-66, 485-488
of unipolar rectangular pulses, 46
of unipolar triangular pulses, 45
Transconductance, 202
Transform(s), bilateral *z*, 527
direct *z*, 524
discrete Walsh, 579
Fourier, 50-57
discrete, 507-509
Hilbert, 128

inverse z, 525
Laplace, 77-79
discrete, 513-514
Transformation, bilinear, 618

Value, mean, 147
mean-square, 147
Variance, 147
Vector, modulation, 104
Voltage, feedback, 211
pumping, 401

Wave, monochromatic, 32
Weak signal, reception and amplification of, 19-20
Width, spectrum, 622

TO THE READER

Mir Publishers welcome your comments on the content, translation and design of the book.

We would also be pleased to receive any suggestions you care to make about our future publications.

Our address is:

USSR, 129820, Moscow, I-110, GSP, Pervy Rizhsky Pereulok, 2, Mir Publishers.

Printed in the Union of Soviet Socialist Republics

Mir Publishers of Mo[illegible] at scientific and technical literature in sixteen languages—En[illegible] French, Italian, Spanish, Portugues[illegible] Czech, Slovak, Serbo-C[illegible], Mongolian, Arabic, Persian (Fars[illegible] Hindi, Bengali and Tamil. [illegible]extbooks for higher technical schoo[illegible] and vocational schools, [illegible]e natural sciences and medicine, i[illegible] cluding textbooks for me[illegible] popular science and science fictio[illegible] The contributors to Mir [illegible]st are leading Soviet scientists and engineers in all fields o[illegible] technology and include more than 40 Members and Corresp[illegible]embers of the USSR Academy of Sciences. Skilled translators [illegible] a high standard of translation from the original Russian. Many [illegible] titles already issued by Mir Publishers have been adopted as texts[illegible] and manuals at educational establishments in France, Switzerlan[illegible]Cuba, Egypt, India, and many other countries. Mir Publishers' books [illegible] foreign languages are exported by V/O "Mezhdunarodnaya Kniga" and can be purchased or ordered through booksellers in your country dealing with V/O "Mezhdunarodnaya Kniga".